INVESTIGACIÓN MATEMÁTICA DE LA EDUCACIÓN
(para personas a las que no les gustan las matemáticas)

Francesc J. Hernàndez

N17

MONOGRAFIES & APROXIMACIONS
Institut de Creativitat i Innovacions Educatives

Monografies & Aproximacions, nº 17

Título: Investigación matemática de la Educación (para personas a las que no les gustan las matemáticas)

Autor: Francesc J. Hernàndez

Universitat de València

Col·lecció dirigida per Rosa Isusi-Fagoaga i Ricard Silvestre Vañó.

Disseny de portada: Silvia Costa

Coordinadora editorial: Rosa Isusi-Fagoaga

ISBN: 978-84-09-18257-2

INVESTIGACIÓN MATEMÁTICA DE LA EDUCACIÓN

(PARA PERSONAS A LAS QUE NO LES GUSTAN LAS MATEMÁTICAS)

Francesc J. Hernàndez

(Universitat de València)

«La línea más corta entre dos puntos

es la línea recta en bajada.»

Jaume Perich

1.- INTRODUCCIÓN. LA CIENCIA, CON LOS OJOS CERRADOS

Los planes de estudio vigentes exigen que el estudiantado realice actividades prácticas y trabajos fin de grado. Posteriormente puede realizar trabajos de máster o tesis doctorales. Respecto del profesorado, se supone que tiene que estar continuamente investigando. Sin embargo, las personas que realizan investigaciones en el ámbito de la educación no suelen tener una buena preparación matemática. Hay varias razones que explican esto: una enseñanza inadecuada del lenguaje matemático, una división del estudiantado en trayectorias de ciencias naturales y trayectorias de ciencias humanas, etc. Por ello, muchas investigaciones educativas carecen de matemáticas o, como mucho, presentan solo unos pocos datos o algunos porcentajes. Esto sería impensable en cualquier otro ámbito de ciencia real. Para ayudar a superar esta situación he escrito este texto, que he elaborado siguiendo tres criterios:

a) Utilizar un lenguaje claro, de manera que el texto pueda ser leído más como una novela que como un árido manual matemático. Y naturalmente eliminar problemas o ejercicios.
b) Utilizar solo investigaciones en las que he participado o ejemplos reales que haya considerado en libros que he escrito o he colaborado a escribir.
c) Referirme a programas informáticos que cualquier persona tenga a su disposición, como son las calculadoras o las hojas de cálculo (solo haré una excepción a este criterio).

También aquí, como decía el Principito, «lo esencial es invisible a los ojos». Porque lo esencial no son las operaciones aritméticas o las figuras geométricas, sino la relación que expresan, lo que ellas nos dicen de la realidad. El grueso manual de Keith Devlin *El lenguaje de la matemática* tiene precisamente ese subtítulo: Hacer visible lo invisible.

Pero las operaciones aritméticas o las figuras geométricas solo nos hablan si sabemos interrogarlas adecuadamente. Por ello, este texto se refiere a la manera matemática de preguntar. En realidad, no hay aquí respuestas, sino cuestiones para aprender a indagar.

Para favorecer la comprensión, se han subrayado algunas palabras. También hay algunos recuadros de ampliación, con una letra más pequeña, y algunas notas al pie de página. Ni las ampliaciones, ni las notas son imprescindibles para la comprensión del texto y se pueden soslayar en una primera lectura. También un par de fragmentos, convenientemente balizados, se pueden saltar.

2.- «ERA UNA CÁLIDA TARDE DE VERANO EN GRECIA...»

Tal vez usted conozca la serie de televisión «Big Bang Theory». En uno de sus episodios (3ª temporada, episodio 10º), Penny le pide a Sheldon que le explique física para impresionar a Leonard. Entonces Sheldon se remonta hasta la antigüedad: «Es una cálida tarde de verano, alrededor del año 600 aC...». Pues a nosotros nos va a pasar lo mismo, a poco que tratemos del lenguaje matemático nos vamos a tener que remontar hasta el mundo griego clásico.

Empezaremos con un ejemplo sencillo: ¿Es usted una persona alta o baja?[1] Si da una contestación es porque está comparando su estatura con la que considera normal, es decir, con aquella estatura que suponemos que es el promedio de las personas... de su edad. Una pregunta tan simple y ya nos encontramos ante dos variables (altura y edad). ¿Dos? Claro que no. Es obvio que tendríamos que considerar también el sexo de usted. ¿Tres variables? Tampoco. Añadamos el grupo étnico... Conclusión: una simple pregunta y ya tenemos varias variables. ¿Varias? Tampoco: tenemos muchísimas. En realidad, casi todo en el universo físico son cosas que varían incesantemente[2]; es decir, cosas que pueden ser así o de otro modo[3].

Tal vez usted haya realizado algún vuelo de larga distancia. Los aviones suelen ascender por encima de los 30.000 pies y entonces, en algunas ocasiones (cuando, por ejemplo, el aparato no realiza giros o no atraviesa turbulencias), los pasajeros pueden tener la sensación que el avión está quieto en el aire. Pero, naturalmente, no lo está[4]. A poco que reflexionen, saben que está en movimiento. Con el ejemplo anterior ha pasado algo así. Parecía una pregunta sencilla la de su altura, pero han comenzado a surgir variables, es decir, cosas que varían. Incluso aquello que nos parecería fijo, también cambia. Tal vez usted formuló su altura en centímetros. Pero el metro es una unidad de medida relativamente reciente. Solo tiene poco más de 200 años. Podríamos utilizar otra unidad de medida (unas líneas atrás hablé de «pies»). Lo mismo podríamos decir de la edad. Hay muchas maneras de medir el tiempo, múltiples calendarios, etc. Escribo estas líneas en el año 2020, pero en el calendario judio es el año 5780, en el islámico el 1441, en el calendario lunar chino el 4716, por no hablar de los mayas, que tienen varios calendarios. Y lo mismo podríamos decir del sexo o del grupo étnico... De golpe, todo se vuelve variable. ¿Todo? Pues bien, volvamos a Sheldon Cooper: En una cálida tarde de verano, alrededor del año 600 aC... algunos griegos descubrieron que no todo era variable,

---

[1] En realidad, sería suficiente con preguntar: «¿Es usted una persona alta?» o «¿Es usted una persona baja?». No hace falta plantear la disyunción.

[2] El filósofo griego Heráclito escribió que no podemos bañarnos dos veces en el mismo río. Luego lo pensó mejor y escribió que ni tan solo podemos bañarnos una vez *en el mismo* río.

[3] En filosofía, algo que puede ser así o de otro modo se denomina «contingente». Lo contrario de «contingente» es «necesario», que es aquello que no puede ser de otro modo.

[4] Otro ejemplo: Usted tiene la sensación que la Tierra está inmóvil, pero...

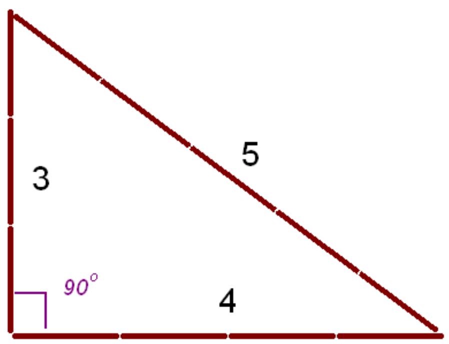
que había cosas que no variaban. Este es el mayor descubrimiento de la historia de la humanidad.

¿Quién hizo ese descubrimiento? En realidad no lo sabemos, pero podemos afirmar que las enseñanzas de Pitágoras ya presuponían ese descubrimiento.

Probablemente el lector de estas líneas asociará Pitágoras con el teorema que lleva su nombre y, tal vez, pueda recordar una formulación clásica, como «en todo triángulo rectángulo, el cuadrado de la hipotenusa es igual a la suma de los cuadrados de los catetos». Dicho así, no parece que nos encontremos ante el mayor descubrimiento de la historia de la humanidad. Da la impresión que el teorema simplemente establece una relación que se da en un triángulo rectángulo entre la «hipotenusa» (palabra griega que significa «opuesto» y que se refiere naturalmente al lado del triángulo opuesto al ángulo recto) y el «cateto» (que, en griego, significa «perpendicular»). Por eso daré otra versión del teorema.

Cojamos un trozo de cuerda o un palo de cualquier longitud. Si formamos un triángulo, cuyos lados midan 3, 4 y 5 veces la longitud de la cuerda o del palo, formaremos una escuadra y dispondremos de un ángulo recto perfecto (la cuarta parte exacta de una circunferencia, dicho de otro modo, un ángulo de 90º) (véase la imagen 1)

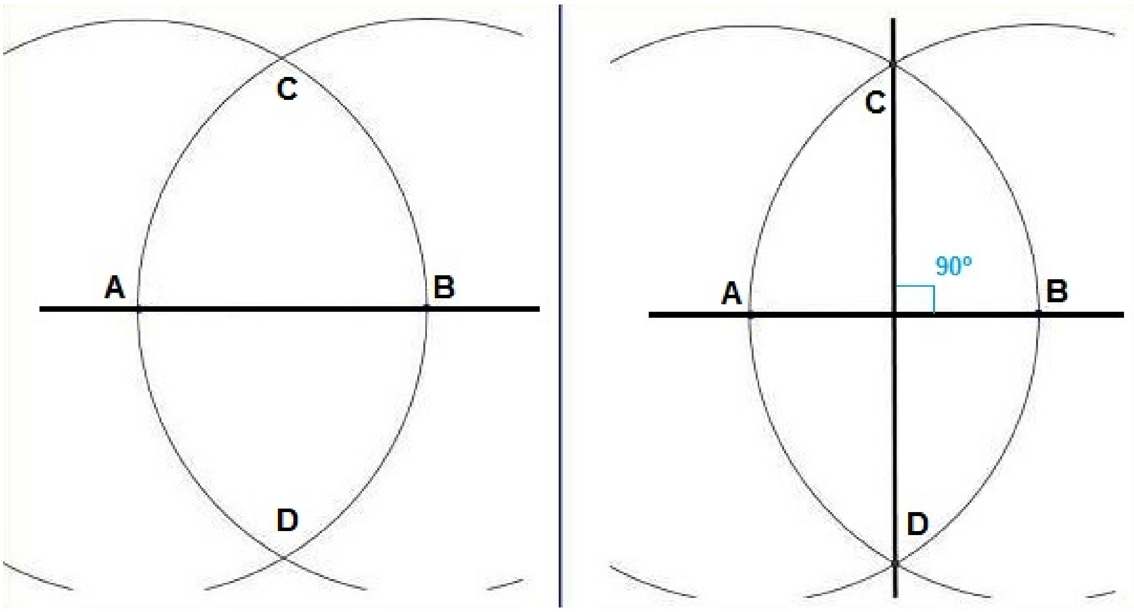
[Imagen 1]



¿Para qué queremos un ángulo recto? Por ejemplo, si tenemos un péndulo en reposo, que define una línea vertical, con la escuadra que hemos formado con el ángulo recto podremos trazar líneas horizontales y luego definir nuevas líneas perpendiculares o paralelas. Y esto es sumamente práctico. Cuando edificamos una vivienda, por ejemplo, disponemos el suelo, los muros y el techo de manera perpendicular, lo que le da la mayor solidez. Esto significa que quien sepa constuir escuadras perfectas, podrá edificar casas más firmes y seguras. De hecho, salvo el Museo Guggenheim de Bilbao (¡que está hecho de titanio!), no conozco otra edificación que no tenga ángulos rectos entre sus muros, el suelo o el techo.

Está claro que podríamos trazar un ángulo recto con métodos meramente geométricos, como puede verse en la imagen 2. Observe la parte de la izquierda de la imagen. Tenemos una línea recta que pasa por los puntos A y B. Desde cualquier punto (por ejemplo, A) trazamos una curva con un compás y obtenemos la curva CBD. Luego desde el punto B, con la misma amplitud, trazamos la curva CAD. Ahora observe la parte de la derecha. Si unimos los puntos C y D con una recta obtenemos una línea perpendicular a la línea que pasa por A y B. Sencillo.

[Imagen 2]



Hay una diferencia muy importante entre el método aritmético (el que hemos visto anteriormente, a partir de los números 3, 4 y 5) y el método geométrico (el que traza curvas). Recuérdese que el trozo de cuerda o el palo que hemos usado antes puede ser de cualquier longitud. Esto significa que también podríamos formar una escuadra, por ejemplo, con una nueva cuerda que midiera el doble que la anterior. Formaríamos ahora un triángulo con los lados de 3, 4 y 5 veces la longitud de la nueva cuerda, pero que, al ser el doble de larga que la anterior, medirían 6, 8 y 10 veces el tamaño de la cuerda primera. Lo mismo podríamos hacer con la mitad o con el triple. Las posibilidades son ilimitadas, pero todas se pueden deducir de la serie: 3, 4 y 5. Eso es constante: el hecho de que esa tríada de números permite formar ángulos rectos. Quien sepa esa relación, sabe algo invariable. Los pitagóricos no sólo encontraron esta relación, sino que, como diríamos ahora, intentaron deducirla, fundamentarla con la razón, y buscar muchas constantes más (en los movimientos astronómicos, en las armonías musicales, etc.).

Probablemente, los sabios egipcios, décadas antes que los griegos, ya tuvieran una cierta idea de esta relación. En el llamado «Papiro de Berlín», encontrado a mitad del siglo XIX, se planteó el problema matemático siguiente, tal vez con finalidad instructiva: Un área de 100 codos cuadrados es igual al área de dos cuadrados más pequeños juntos. El lado de uno es igual a ½ + ¼ del lado del otro (los egipcios empleaban sólo fracciones con numerador 1 y excepcionalmente ⅔). ¿Qué lado tienen? La respuesta es lógicamente 8 y 6 codos (6 codos son ½ + ¼ de 8 codos), lo que nos lleva a la relación $10^2 = 8^2 + 6^2$ ($100 = 64 + 36$), esto es, a la relación de un triángulo rectángulo de hipotenusa 10 y catetos 8 y 6 , que tiene la misma forma que el triángulo de lados 5, 4 y 3.

Estos tres números (3, 4 y 5) forman una «terna pitagórica», es decir, generan un triángulo rectángulo. Hay más, como, por ejemplo: 5, 12 y 13; 7, 24 y 25; 8, 15 y 17; 9, 40 y 41; 11, 60 y 61; 12, 35 y 37; 13, 84 y 85; 16, 63 y 65; 20, 21 y 29; 28, 45 y 53; 33, 56 y 65; 36, 77 y 85; 39, 80 y 89; 48, 55 y 73, 65, 72 y 97, etc.

En la *Geometría* de Euclides se proporciona una fórmula para generar infinitas ternas posibles: Si m y n son dos números naturales ($m > n$),[5] y a la hipotenusa y b y c los catetos de un triángulo rectángulo, las ternas se pueden formar con las fórmulas:

$$a = m^2 + n^2;\ b = 2mn;\ c = m^2 - n^2$$

También podemos formar ternas pitagóricas con la serie de Fibonacci. La serie de Fibonacci comienza con 0, 1 y a partir del miembro 3º es la suma de los dos anteriores. así:

0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, ...

Podemos formar ternas pitagóricas usando cuatro números consecutivos (cuaternaria) de la serie de Fibonacci. Un cateto es igual a la multiplicación de los extremos de la cuaternaria; el otro es igual a la duplicación del producto de los dos números centrales. La hipotenusa es, lógicamente, la suma de los cuadrados de los catetos. Por ejemplo, tomamos (13, 21, 34, 55), entonces:

$$b = 13 \,.\, 55 = 715;\ c = 2(21 \,.\, 34) = 1428;\ \text{y } a = 715^2 + 1428^2 = 1597^2.$$

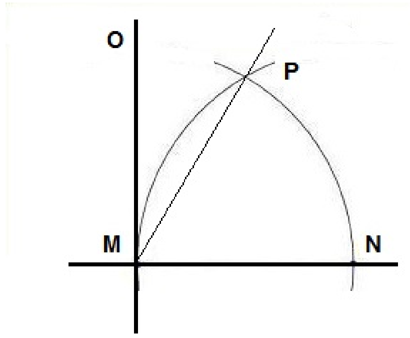
Por lo tanto, 715, 1428 y 1597 forman una terna pitagórica.

3.- NO SOLO EDIFICIOS SÓLIDOS, SINO LA CÚPULA CELESTE

Antes comenté que con ángulos rectos podrían construirse edificios más sólidos. Es cierto, pero no solamente sirven para esto. Pondré otro ejemplo. Con un ángulo recto es muy fácil construir un ángulo de 30º y esto fue muy importante. Explicaré el procedimiento, representado en la imagen 3.

[5] El signo «>» significa mayor que. En este texto se usarán los signos matemáticos habituales.

[Imagen 3]



Supongamos que tenemos dos líneas perpendiculares, la que pasa por los puntos MN y la que pasa por los puntos MO. Desde el punto M trazamos la curva NP (con un compás, una cuerda) y, con la misma amplitud (de compás), trazamos la curva MP desde el punto N. Unimos los puntos MP. Esa línea, la que pasa por los puntos MP, forma un ángulo de 30º con la línea MO. ¿Para qué queremos tener un ángulo de 30º? Le puede parecer algo insignificante, pero no lo es. Piénsese que los seres humanos comenzaron a orientarse a partir de la bóveda celeste, que dividieron en doce fragmentos, cada uno de los cuales mide 30º (360º/12=30º). Después imaginaron que las estrellas de cada fragmento formaban figuras, les pusieron nombres –por cierto, bastante imaginativos– y así era más fácil reconocerlas. De este modo nacieron las constelaciones y los signos zodiacales, y se puede navegar por el mar y engañar a la gente con los horóscopos.

Con los ángulos rectos también pueden construirse triángulos rectángulos, que permitieron el desarrollo de la trigonometría, pero explicar esto excede de la pretensión de este libro (aunque en un epígrafe posterior utilizaremos alguna noción trigonométrica, pero no se preocupe porque allí se explicará).[6]

Adviértase que los pitagóricos, por así decir, construyeron un puente desde la aritmética (números) hasta la geometría (triángulos, formas) y lo atravesaron por primera vez. De este modo encontraron relaciones generales y, en cierto modo, necesarias. Es lo que Platón denominó «ciencia» (*episteme*) y Aristóteles «sabiduría» (*sofía*). Es decir, se trata de buscar relaciones constantes en las ilimitadas variables que nos ofrece la realidad, y hacerlo con el lenguaje matemático. Recuerde: eso es la ciencia.

---

[6] Algo parecido a lo explicado de Pitágoras, se podría decir del Teorema de Tales de Mileto. Pero dejo esta ampliación para la curiosidad del lector o la lectora.

4.- ¿En Malta beben mucha Coca-cola?

Las investigaciones suelen comenzar haciendo un diseño general (que muchas veces se modifica a lo largo del proceso). Antes de entrar en el lenguaje matemático (que es de lo que trata este texto), hay que recolectar una serie de datos. Pues bien, muchos errores en las investigaciones proceden de una recolección de datos poco escrupulosa. Pondré un ejemplo real.

El prestigioso y riguroso periódico británico *The Times* publicó en agosto de 2011 un reportaje sobre el consumo de Coca-cola, con un gráfico a doble página donde se representaba el consumo anual per cápita de esa bebida en diversos países europeos. Tal vez por la afición británica a su sistema de medidas imperiales, la cantidad de líquido consumida no se expresaba en litros, sino en el número equivalente de los clásicos botellines de Coca-cola de 6,5 onzas. Según el gráfico, en Estados Unidos el promedio era de 394 botellines por persona y año. En Malta el promedio era de 606 botellines. ¿Por qué los habitantes de la isla mediterránea consumían un 54% más de refresco que los habitantes del país donde Pemberton inventó esa bebida?[7] ¿A qué se debe el incremento que hace de los malteses líderes mundiales del consumo del líquido refrescante según *The Times*?

Es facil suponer que no se trata de ninguna pasión maltesa por la Coca-cola, sino simplemente de un cálculo erróneo, probablemente derivado de una confusión en las personas que habían cubierto la información, ya sean periodistas o responsables del grafismo, en el dato de la población de hecho y la población de derecho de la isla de Malta. Si en lugar de considerar las personas que efectivamente se encuentran en la isla (población de hecho), muchas de ellas turistas, se considera la cantidad de la ciudadanía maltesa (población de derecho), la cantidad de botellines de Coca-cola por persona naturalmente se incrementa.

5.- Un ejemplo para empezar a calcular

De estas explicaciones previas debe recordar que buscamos algo constante. Si no lo obtenemos, como ya averiguaron los antiguos griegos, simplemente no estamos haciendo ciencia. Comencemos con un ejemplo. Observemos la tabla siguiente (tabla 1)

[7] El incremento de un 53% es fácil de calcular. Se puede hacer una regla de tres: 394 botellines és a 100, como 606 es a *x*. El valor de *x* es de 153,8. Si restamos 100, obtenemos el incremento del 53,8%, redondeando 54%.

Tabla 1. Cuotas medias en los centros de privados en enseñanzas universitarias y no universitarias en España (euros)

Centros privados	1999/00	2004/05	2009/10	2014/15
Enseñanzas universitarias	3.291 €	4.064 €	4.718 €	5.411 €
Enseñanzas no universitarias	819 €	1.136 €	1.541 €	1.770 €

Fuente: Instituto Nacional de Estadística. Encuesta de Financiación de la Enseñanza Privada 1999/2000, 2004/2005, 2009/2010 y 2014/2015 (www.ine.es).

La tabla 1[8] recoge las cuota media que pagan los estudiantes de centros privados de enseñanzas universitarias y de enseñanzas no universitarias en España en una serie de cursos[9] en los que el Instituto Nacional de Estadística (INE) realiza esta encuesta. Los datos están expresados en euros.

Aquí tenemos unas variables. En realidad, tres variables (cuota de enseñanza privada universitaria, cuota de enseñanza privada no universitaria y cursos o años). La pregunta clave es ¿qué hay de constante? Aparentemente nada. Si añadimos esa tabla a un trabajo de investigación realmente no estamos investigando nada. Informamos de cómo varían las variables, pero eso no es hacer ciencia, sino ofrecer una mera descripción. Y naturalmente tampoco hacemos ciencia si ponemos la tabla y nos dedicamos a glosarla, diciendo lo mismo que aparece en ella. Este es un buen recurso para incrementar las páginas de nuestro trabajo, pero un mal recurso para penetrar en el ámbito de la ciencia.

Analicemos la tabla con más detalle. Nos dice que, por ejemplo, un estudiante que se matriculó en un centro privado de enseñanzas universitarias pagó, como promedio, 3.291 € en el curso 1999/2000 y 5.411 € en el curso 2014/2015. Entonces nos podemos preguntar: ¿Qué se pagó en el curso 2007/2008 o 2012/2013? La tabla no nos lo dice, y menos aún qué se pagaría en el 2019/2020 o en el 2024/2025. Pero entonces podemos establecer un modelo (más adelante trataré con más detalle qué significa un «modelo»).

Vamos a suponer que los incrementos anuales son los mismos, es decir, que cada año la tasa de incremento de las cuotas de matrícula fue idéntica. Si fuera la misma, ya tendríamos un valor constante. Y si disponemos de ese valor, podremos calcular las cuotas de los otros cursos que he indicado. Para simplificar el cálculo, nos referiremos al primer año de cada curso (por ejemplo, en lugar de 1999/2000 tomaremos 1999)[10] y solo los valores extremos, esto es el valor inicial y el valor

[8] Puede descargar las tablas de este libro de la dirección: http://www.uv.es/fjhernan/investigando.xlsx

[9] Recuérdese que, en el hemisferio Norte, los cursos comienzan después del verano y concluyen antes de las vacaciones de verano del año siguiente, por lo que se aluden con dos años naturales. En el hemisferio Sur, sin embargo, el verano viene a coincidir con el comienzo del año, por lo que los cursos se designan con un solo año natural.

[10] Daría lo mismo si tomáramos el segundo año, el 2000.

final. De esta manera podemos formar la siguiente tabla 2 que simplifica la anterior.

Tabla 2. Simplificación de la tabla 1

	Valor inicial (1999)	Valor final (2014)
Cuotas universitarias	3.291 €	5.411 €
Cuotas no universitarias	819 €	1.770 €

Fuente: Elaboración propia de la tabla 1.

Si tenemos un valor inicial ($V_{inicial}$) y un valor final (V_{final}) y unos años (*años*), la tasa de incremento anual (*tasa*) corresponde a la fórmula siguiente[11]:

$$V_{final} = V_{inicial}\,(1 + tasa)^{años}$$

Esta es una fórmula muy conocida en cálculos económicos y financieros. Entonces, despejando la fórmula para calcular la tasa, tendremos:

$$tasa = \sqrt[\text{años}]{\frac{V_{final}}{V_{inicial}}} - 1$$

El número de años es de 15 ($2014 - 1999 = 15$), por lo que para el caso de las cuotas de enseñanzas universitarias:

$$tasa = \sqrt[15]{\frac{5.411}{3.291}} - 1 = 0{,}0337 = 3{,}37\%$$

Y para el caso de las cuotas de enseñanzas no universitarias:

[11] La deducción de esta fórmula es sencilla. Al final del año 1, el $V_{final} = V_{inicial} + (V_{inicial} \times tasa) = V_{inicial}(1 + tasa)$; al final del año 2, será: $V_{final} = [V_{inicial}(1 + tasa)] + [V_{inicial}(1 + tasa) \times tasa] = V_{inicial} + (V_{inicial} \times tasa) + (V_{inicial} \times tasa) + (V_{inicial} \times tasa^2) = V_{inicial}(1 + tasa)^2$, y así sucesivamente; se trata de la misma fórmula en la que el exponente del último factor coincide con los años.

$$tasa = \sqrt[15]{\frac{1.770}{819}} - 1 = 0{,}0527 = 5{,}27\%$$

Aunque es probable que usted sepa utilizar la calculadora, no está de más explicar cómo realizar estas operaciones de modo sencillo.

Hacer una raíz con el índice n (en este caso: 15) de un número es lo mismo que elevar ese número a $\frac{1}{n}$ (en este caso: $\frac{1}{15}$). Por ello, lo mejor es comenzar con la calculadora dividiendo 1 por 15 y guardando ese valor en la memoria. Después hacemos la división y la elevamos a la cantidad que tenemos guardada en la memoria (mediante la tecla [x^y] –aquí utilizaré los paréntesis cuadrados para indicar simplemente una tecla de la calculadora; por ejemplo: [1] es presionar la tecla del «1»–) y luego restamos la unidad. En el primer ejemplo, la serie completa de teclas y los resultados parciales que aparecen en la pantalla (en un recuadro)[12] serían:

[1] [/] [1][5] [=]

0,06666666666666666666666666666667

[M+]

[5][4][1][1] [/] [3][2][9][1] [=]

1,6441810999696140990580370707991

[x^y] [MR] [=]

1,0337050630622411241083276652991

[-] [1] [=]

0,03370506306224112410832766529915

Si quisiéramos expresarlo porcentualmente, solo tenemos que dividir esta cantidad por 100.

[/] [1][0][0] [=]

3,370506306224112410832766529915

[12] El número de decimales puede variar en función del programa que usted utilice.

Es decir: 3,37%

Para calcular el segundo ejemplo, podemos saltar el primer paso, porque ya tenemos el valor $\frac{1}{15}$ en la memoria.

Un <u>ERROR habitual</u> es calcular el porcentaje de incremento entre el valor inicial y el final y dividirlo por el número de años. En este caso el procedimiento <u>erróneo</u> sería:

Si	3.291 € corresponde al	100%
Entonces	5.411 € corresponde al	x

Por lo que (según la clásica «regla de tres»):

$$x = \frac{5.411 \times 100}{3.291} = 164{,}418$$

Luego se supone que se habrá producido un incremento del 64,418%, que dividido por los 15 años dará un incremento anual igual a 4,29%. El resultado es lógicamente <u>erróneo</u>, porque, como vimos más arriba, el resultado correcto es 3,37%.

Si sabemos cuál es el valor de la tasa anual podemos enfrentarnos a cuestiones como averiguar si los valores intermedios (curso 2004/2005 y 2009/2010) están por encima o por debajo de los que corresponderían con la tasa constante o podemos hacer una estimación de las cuotas correspondientes a otros años o incluso una previsión de valores futuros (más adelante ofreceré un método más sencillo para averiguar esto mismo). Incluso podemos utilizar como valor inicial 1999 (es decir: 1999/2000) y como valor final 2009 y hacer una estimación de cuál tendría que ser el valor para 2014 y luego compararlo con el valor real.

Adviértase además que, en el caso de este sencillo ejemplo, la tasa de las cuotas de enseñanzas no universitarias es superior a la de enseñanzas universitarias a pesar de que en términos absolutos el valor de las cuotas es menor. Esta sería una conclusión muy interesante de la investigación, pero que depende, lógicamente, del modelo que hemos <u>elegido</u> (a saber: que las tasas serían iguales en todos los años).

6.- OTRO EJEMPLO PARA EMPEZAR A DUDAR

Tal vez algún lector se haya sorprendido de que hable del modelo que hemos «elegido», como si por el hecho de usar un lenguaje matemático no hubiera ya ningún resquicio a la elección. Incluso alguién podría invocar al mismísimo Isaac Newton, una de cuyas frases más célebres es: *Hypotheses non fingo*, «No compongo hipótesis» (es decir, no hago suposiciones). Naturalmente, esta frase (¡y todas!) tienen que interpretarse en su contexto, pero para entender mejor que el lenguaje matemático no está reñido con algunas suposiciones (inherentes al mismo uso de modelos), veamos un curioso ejemplo.

Imaginemos una <u>suma infinita</u> (que llamaremos S) con la forma:

$S = 1 - 1 + 1 - 1 + 1 - 1 + 1 - 1 + 1 \ldots$

¿Qué vale S, es decir, qué suma esta serie ilimitada? En principio, podríamos agrupar los números de la serie por parejas:

$+ 1 - 1; + 1 - 1; + 1 - 1 \ldots$

Como + 1 - 1 es cero, entonces:

$S = 0 + 0 + 0 \ldots$

Por lo que podríamos concluir que $S = 0$. Ahora bien, también podríamos hacer los agrupamientos a partir del segundo sumando, porque también – 1+ 1 es cero. Así la parejas:

$- 1 + 1; - 1 + 1; - 1 + 1$, etc.

Serían equivalentes a una serie de ceros y entonces podríamos reformular la suma como:

$S = 1 + 0 + 0 \ldots$

Entonces, llegaríamos a la conclusión de que $S = 1$. ¿En qué quedamos, $S = 0$ o $S = 1$? Pues bien, todavía sería posible encontrar <u>otra</u> solución. La suma, a partir del segundo sumando:

$- 1 + 1 - 1 + 1 - 1 + 1 - 1 + 1 \ldots$

Es la inversa de:

$+ 1 - 1 + 1 - 1 + 1 - 1 + 1 - 1 \ldots$

Que es precisamente S. Por lo tanto, a partir del segundo sumando podemos sustituir la serie por $-S$, de forma que:

$S = 1 - S$ y pasando la S de la derecha al otro miembro de la ecuación:

$2S = 1$ y por tanto, dividiendo ambos miembros por 2:

$S = \frac{1}{2}$

Entonces, ¿cuál es el valor de S? ¿0?, ¿1?, ¿$\frac{1}{2}$?

En realidad, lo que hemos hecho en cada uno de los tres casos es tomar una serie de decisiones (sumar dos a dos, sumar dos a dos a partir del segundo sumando o considerar una parte de la serie el inverso de *S*), que dan resultados distintos [13]

7.- La línea más próxima a los puntos

Volvamos al ejemplo del epígrafe anterior, sobre las cuotas que se pagaban en los centros de enseñanza privados. Volveré a copiar la tabla 1, para que no tenga que ir pasando páginas adelante y atrás.

Tabla 1. Cuotas medias en los centros de privados en enseñanzas universitarias y no universitarias en España (euros)

Centros privados	1999/00	2004/05	2009/10	2014/15
Enseñanzas universitarias	3.291 €	4.064 €	4.718 €	5.411 €
Enseñanzas no universitarias	819 €	1.136 €	1.541 €	1.770 €

Fuente: Instituto Nacional de Estadística. Encuesta de Financiación de la Enseñanza Privada 1999/2000, 2004/2005, 2009/2010 y 2014/2015 (www.ine.es).

Vamos a proceder a realizar un gráfico con la tabla 1, es decir, representar los valores en unos ejes cartesianos (o unos ejes XY).

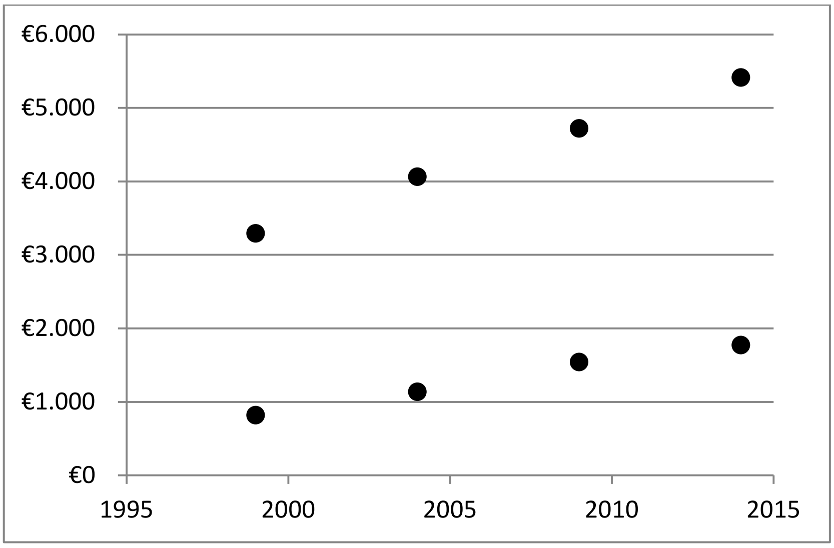
Se cuenta que René Descartes inventó el diagrama que lleva su nombre (el latín su nombre era «*Cartesius*» de donde procede «cartesiano») viendo cómo unas moscas deambulaban por su mesa. Los desplazamientos de los insectos se podían describir –penso él– a partir de las distancias que mantenían los bichos con los bordes de la mesa. Sea verdad o no, lo cierto es que la mesa y las moscas se parecen a los ejes cartesianos, que son dos líneas rectas perpendiculares (que forman un ángulo recto de 90º: ¡volvemos a los pitagóricos! ¡qué importantes son los ángulos rectos!), que determinan un plano en el que se pueden trazar puntos.

Podemos hacerlo manualmente o mecánicamente (que siempre resulta mejor). Se cargan los datos de la tabla 1 en una hoja de cálculo (yo he utilizado el programa Excel, pero sirve cualquier otro) y después se seleccionan las celdas con los datos

[13] Para expresar la multiplicación de dos factores, en este texto se utilizará el signo «×», el punto «.» o simplemente ubicar los factores juntos. Así, será lo mismo «a×b», «a*b», «a.b», «ab» o, lógicamente, «ba».

pasando el cursor por ellas mientras se mantiene presionado el botón izquierdo del ratón y luego, en los botones de instrucciones superiores, se pide al programa que inserte un «gráfico de dispersión XY» (o «de burbujas»). Se obtiene así el Gráfico 1.

[Gráfico 1]



¡Aquí tiene las moscas y la mesa de Descartes!

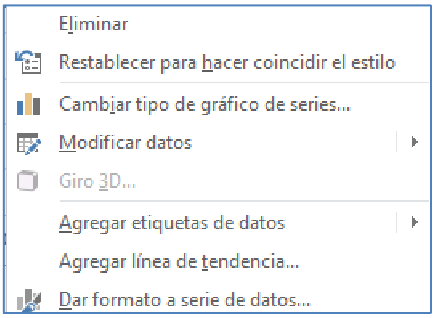
Los cuatro puntos superiores corresponden a las cuotas universitarias y los cuatro inferiores a las cuotas no universitarias. Dejemos de lado, por el momento, los cuatro puntos inferiores y concentrémonos en los cuatro superiores. Está claro que en ese plano podríamos trazar infinitas líneas rectas (usted podría coger una regla y un bolígrafo y dedicarse a rayar el rectángulo donde está el gráfico hasta cubrirlo totalmente: ¡infinitas líneas rectas!). Pues bien, de todas ellas habrá <u>una recta</u> que será la que tendrá menores distancias a los cuatro puntos superiores[14]. Esta línea (en este caso, <u>recta</u> –más adelante veremos líneas curvas–) recibe el nombre de <u>línea de tendencia (recta o lineal)</u>.

Sería una pérdida de tiempo que le explicara a usted cómo puede trazar manualmente esa línea de tendencia, cuando disponemos de ordenadores y

[14] La distancia entre una línea se mide con la perpendicular a la línea que pasa por el punto.

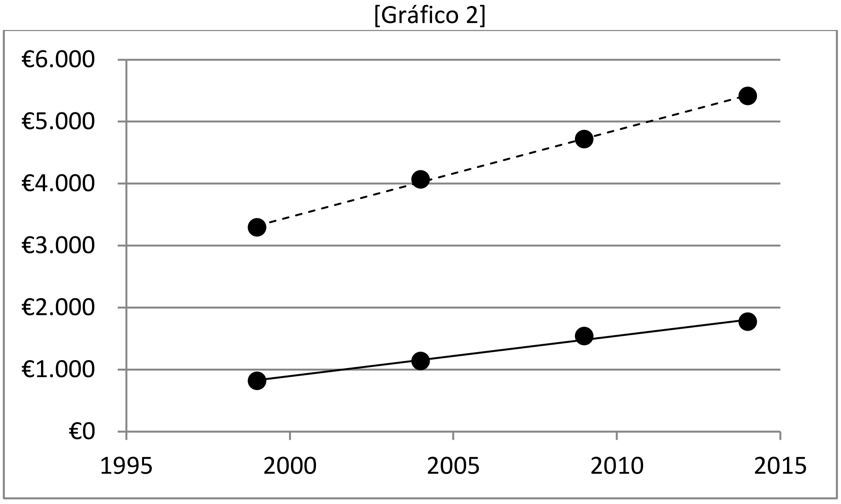
programas que lo hacen en una fracción de segundo. Por ello, utilizaremos estos «amigos inhumanos»[15]. El procedimientos es sencillo. Realizado el gráfico en la hoja de cálculo (o incluso, trasladado a un editor de textos), simplemente se coloca el cursor sobre un punto cualquiera de la serie (los cuatro puntos superiores, en este caso) y, haciendo clic en el botón derecho del ratón (o su equivalente), se abre una ventana de diálogo y se le pide al programa <Agregar línea de tendencia> (véase la imagen 4, penúltima línea).

[Imagen 4]



Después podemos hacer lo mismo con los cuatro puntos inferiores. Los programas nos permiten definir las características de las líneas (grosor, tipo, incluso si queremos que se amplíen por detras –hacia la izquierda– o por delante –hacia la derecha–). De este modo obtenemos el gráfico 2.

[Gráfico 2]



[Para distinguir las dos líneas he trazado las líneas de manera diferente: Cuotas universitarias: línea discontinua superior; cuotas no universitarias: línea contínua inferior]

[15] Tomo esta agradable expresión del periodista, especializado en ajedrez, Leontxo García.

El gráfico 2 nos permite observar a simple vista que, por ejemplo, el punto de 2004 (curso 2004/2005) en el caso de las cuotas universitarias o el punto de 2008 (2009/2010) en el curso están ligeramente por encima de la línea (¿Lo ve?). ¿Qué significa esto? Pues algo bastante sencillo: si calculamos el valor correspondiente para esos puntos con la fórmula citada anteriormente y con los valores de las tasas establecidos anteriormente veremos que el valor estimado queda por debajo del valor real que aparece en la tabla 1. Más adelante ofreceré un procedimiento más sencillo.[16]

[16] Puede ser que usted se pregunte: «Y si hay un procedimiento más sencillo, ¿por qué no lo explica ahora». Escribí al principio que debe leer este libro como una novela.

8.- CRUZANDO DE NUEVO EL PUENTE, EN EL OTRO SENTIDO

Y ahora vamos a proceder a realizar una cosa muy importante, así que me gustaría que prestara la mayor atención al argumento.

Recordará usted que, con los pitagóricos, atravesamos el puente que iba desde los números (el ejemplo de 3, 4 y 5) hasta las formas (el triángulo rectángulo). Pues ahora vamos a cruzar de nuevo el puente, pero en sentido contrario: vamos a pasar de las formas a los números, por así decir. Lo podemos hacer porque, en el plano cartesiano, toda recta está asociada a una función del tipo:[17]

$$y = ax + b$$

Que por ello se suele denominar función de la recta (en realidad, toda línea está asociada a una función, pero de esto hablaremos más adelante).

Recuerde que una función o aplicación es una regla que relaciona un elemento de un conjunto con un elemento de otro conjunto. La expresión $y = f(x)$ significa precisamente eso, que podemos entender la variable y como una función (f) de la variable x, esto es como $f(x)$.

La expresión $y = f(x)$ es una ecuación, es decir, una igualdad que expresamos con el signo «=», que fue utilizado por primera vez por Robert Recorde en el año 1557, y lo que hizo fue precisamente expresar algo, la igualdad de dos cosas, mediante dos líneas rectas iguales (¡y de eso precisamente estamos hablando todo el tiempo, de expresar relaciones mediante líneas y viceversa!).

Atendamos al significado de a y b en la función de la recta. Por un lado, b es naturalmente el valor que tiene y cuando x es cero (si $x = 0$, entonces $y = ax + b$, se convierte en: $y = a0 + b = b$), es decir, el punto en el que se cortan la línea de tendencia y el eje y. Por otro lado, a indica la pendiente de la recta, esto es, lo inclinada que esté.

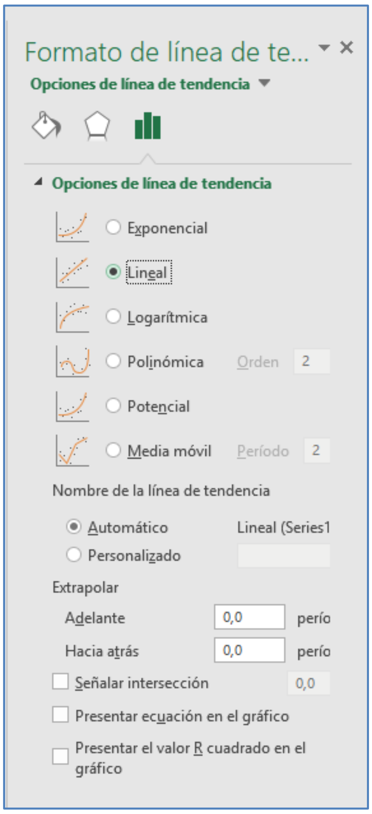
Como veremos más adelante, en realidad, a es equivalente a la tangente del ángulo que forman la línea de tendencia y el eje x (o cualquiera de sus paralelas[18])

[17] En algunos manuales se usan las letras «m» y «n»; entonces la forma de la función de la recta sería: $y = mx + n$.

[18] Por el teorema de Tales, ya mencionado, que trata de la intersección de una paralelas por una línea.

La ecuación de la línea de tendencia puede establecerse manualmente o mecánicamente (lo que resulta más fácil). Al agregar la línea de tendencia (véase la anterior imagen 4), el programa nos ofrece la posibilidad de <Presentar ecuación en el gráfico>, como aparece en la imagen 5 (observe la penúltima línea de la ventana de diálogo).

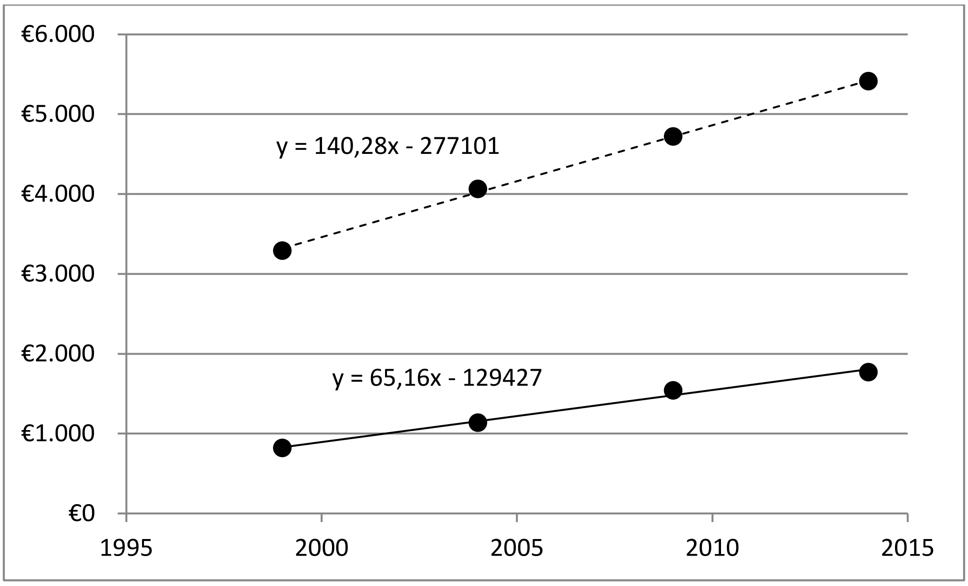
[Imagen 5]



(Las imágenes 4 y 5 corresponden al programa Excel. Es lógico que otros programas presentan ventanas de diálogo semejantes, o que en otras versiones –anteriores o posteriores–, la forma será ligeramente diferente)

El resultado que podemos obtener se presenta en el gráfico 3

[Gráfico 3]



[Cuotas universitarias: línea discontinua; cuotas no universitarias: línea contínua]

Disponemos, pues, de dos procedimientos para hacer estimaciones, en cada caso. Por ejemplo, podemos calcular la cuota que correspondería a un determinado curso (año), con la fórmula general del epígrafe 5[19] y las tasas correspondientes que hemos calculado anteriormente o con la función que hemos establecido ahora (que tal vez sea un procedimiento más sencillo).

Naturalmente, si un punto se aleja mucho de la línea de tendencia puede ser por causas espúreas (tal vez una mala medición o alguna característica peculiar de ese dato). Por eso siempre es recomendable visualizar los valores de una tabla en unos ejes cartesianos (en una gráfica).

Ahora se entenderá que la tasa o la ecuación nos acercan a esa aspiración a descubrir algo constante, a ese intento de huir de la inmensa variabilidad de lo real, que anima a la ciencia desde los pitagóricos. Repito: las tablas no son ciencia (y su comentario, menos). La ciencia comienza cuando seguimos nuestro camino buscando relaciones permanentes, constantes, a las que las tasas o las funciones nos pueden aproximar.

[19] Recuerde: $V_{final} = V_{inicial}\,(1 + tasa)^{años}$.

9.- ¡CUÁNTOS CUANTILES!

Para comprender el epígrafe siguiente –y algunos otros posteriores– he de introducir una explicación sobre qué son los cuantiles.

Tal vez usted no escuchó nunca la palabra «cuantil», pero sí «percentil». Por ejemplo, es habitual que los ginecólogos o los pediatras informen a los progenitores de que los recien nacidos están en tal o cual percentil en su desarrollo. Pues bien, los percentiles son un tipo de cuantiles.

Supongamos que asociamos a los individuos de un grupo con una variable (por ejemplo, la altura, que era nuestro primer ejemplo) y después los ordenamos de menor a mayor según el valor de la variable (los valores menores, a la izquieda). Tenemos entonces una muestra ordenada. Diremos que una determinada altura es, por ejemplo, el cuantil 0,40 si el individuo que tiene esa altura tiene a su izquierda el 40% de la muestra (es decir, que el 40% de la muestra presenta valores inferiores). Se podría decir que un cuantil es como un corte que hacemos en una muestra ordenada.

Los cuantiles más usados son los cuartiles, que dividen la muestra en cuatro partes, de manera que el cuartil primero tiene el 25% de la muestra a su izquierda, el cuartil segundo el 50%, el cuartil tercero el 75% y el cuartil cuarto el 100%; los quintiles que dividen la muestra en cinco partes, cada una con el 20% del total; los deciles que la dividen en diez partes, representando cada una el 10% del total, y los centiles o percentiles (ya mencionados antes) que dividen la muestra en cien partes, cada una con el 1% del total. Es obvio que el cuartil segundo corresponde al decil quinto y al percentil quincuagésimo, y todos ellos equivalen a la mediana (el valor que está en el punto medio) de la distribución de la muestra.

En algunas investigaciones se puede distribuir la muestra en cuantiles, quintiles o deciles, y calcular, por ejemplo, el promedio de cada grupo. Esto permite apreciar mejor la diferencia entre los individuos. De esta manera, por ejemplo, la Unión Europea calcula la desigualdad de los diversos países, como la proporción entre los ingresos anuales del quintil superior y la del inferior. Así, podemos averiguar qué paises son más igualitarios y en qué países hay más desigualdad (como es el caso de España).

Así pues, una de las cosas interesantes que se puede hacer en la investigación educativa cuando se dispone de una muestra es ordenarla, calcular cuantiles y establecer relaciones entre ellos. Obtendremos así medidas de desigualdad no solo social, sino también de recursos, rendimientos, etc. (más adelante ofreceremos otros procedimientos para considerar la desigualdad en la investigación educativa).

10.- Pendientes y ángulos

Como es conocido, la Organización para la Cooperación y el Desarrollo Económico (OCDE) realiza desde años las denominadas pruebas PISA (acrónimo de «Programme for International Student Assessment»). Se trata de una prueba a una muestra de estudiantes del curso correspondiente a los 15 años en las áreas de matemáticas, ciencias naturales y lengua (de 15 años como moda estadística). Se ofrecen tanto los resultados generales, en una escala de puntuación propia (que oscila en torno a los 500 puntos), como también los porcentajes de estudiantes en cada nivel de rendimiento. En este sentido, se distingue lo que PISA considera «bajo rendimiento» (los dos niveles inferiores de los cinco que se establecen). Resulta relevante la noción de bajo rendimiento por cuanto la Unión Europea ha establecido en su Estrategia Educación y Formación 2020 el objetivo de que tal indicador se sitúe por debajo del 15% para el final de la década en los distintos países de la Unión Europea. PISA ofrece sus resultados desagregados por múltiples variables, una de las cuales es la que denominan «estatus socioeconómico» (ESCS, por las siglas utilizadas). En la base de datos de los resultados de PISA entontramos el porcentaje de bajo rendimiento correspondiente a cada cuartil de estatus socioeconómico para el caso de la prueba de ciencias naturales. (En el epígrafe anterior ya se explicó lo que era un cuartil).

Pues bien, según los datos de PISA para el conjunto de la OCDE, los datos de bajo rendimiento en ciencias naturales según los cuartiles del índice de estatus económico son los recogidos en la tabla 3, donde también se ha añadido la fila del «no bajo rendimiento», esto es, el valor complementario al bajo rendimiento, es decir, el porcentaje de estudiantes que en cada cuartil de estatus socioeconómico no presentan bajo rendimiento, lo que se obtiene fácilmente restando cada porcentaje de la unidad.

[Tabla 3]

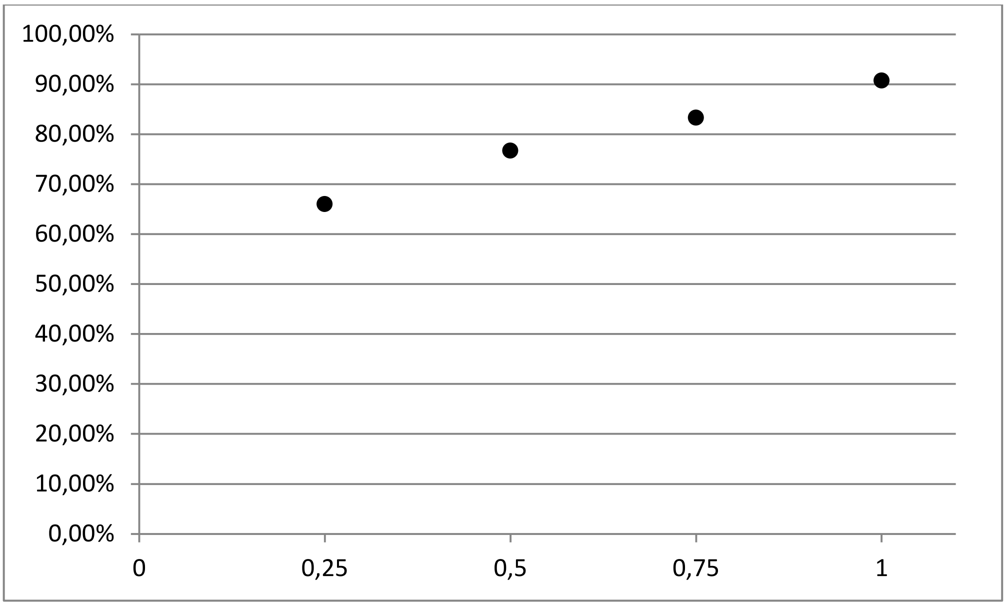
	Cuartil 1º	Cuartil 2º	Cuartil 3º	Cuartil 4º
Porcentaje de bajo rendimiento	34,0%	23,3%	16,7%	9,3%
Porcentaje de no bajo rendimiento	66,0%	76,7%	83,3%	90,7%

Fuente: Elaboración propia datos PISA (OCDE)

Lo que nos dice esta tabla 3 es que en el caso de los estudiantes con estatus socioeconómico inferior (cuartil 1º) una tercera parte aproximadamente tienen un bajo rendimiento educativo (o, a la inversa, dos terceras partes no lo tienen) y que solo menos de una décima parte de los estudiantes con estatus socioeconómico superior (cuartil 4º) tienen bajo rendimiento educativo (a la inversa, un 90% de estos estudiantes no lo tienen). Es trivial que si estos porcentajes fueran iguales el estatus socioeconómico no tendría relación con el rendimiento educativo, lo que naturalmente no es el caso.

Recuérdese que este ejemplo pretendía ilustrar otro uso de la línea de tendencia. A continuación vamos a representar los valores de no bajo rendimiento de la tabla 3 en los cuantiles del índice de estatus socioeconómico, que representaremos en el eje x. Se forma así el gráfico 4.

[Gráfico 4]



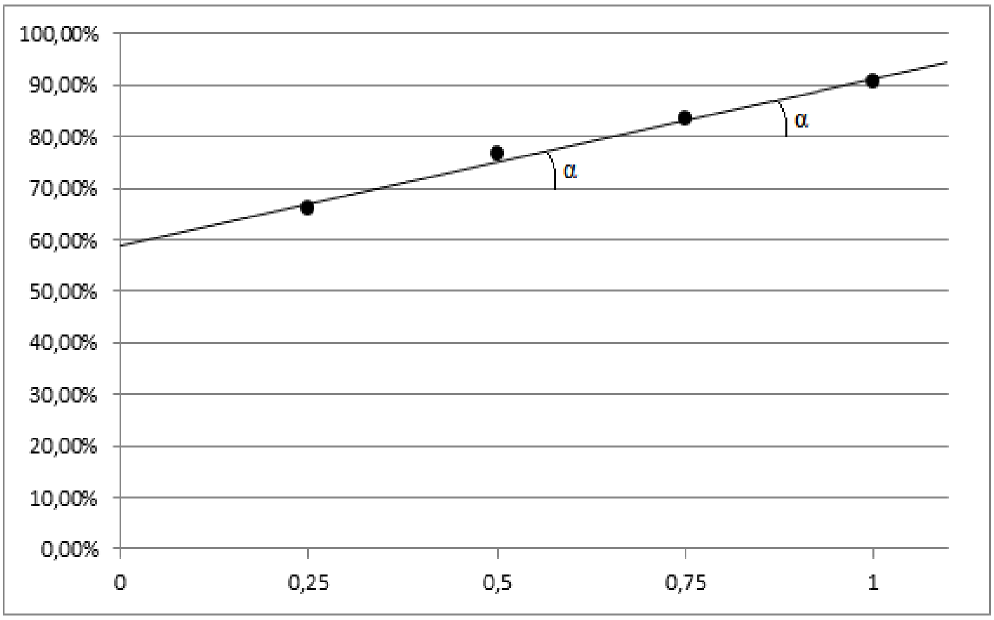
Fuente: PISA 2015 Results (Volume I), OECD Publishing, Paris. DOI: http://dx.doi.org/10.1787/9789264266490-table124-en Tabla I.6.6a. Valores globales OCDE

A continuación, añadiremos una línea de tendencia recta. Esa línea de tendencia corta las líneas paralelas horizontales del gráfico, determinando siempre los mismos ángulos. También hemos marcado ese ángulo que se repite con un pequeño arco y lo designaremos con la letra griega α. Ese ángulo representa la pendiente de la línea de tendencia recta.

En este caso hemos utilizado una línea de tendencia recta, que también presenta una función que, al tratarse de una recta, se corresponde con la ecuación que tiene la forma general:

$$y = ax + b$$

[Gráfico 5]



En el ejemplo, la función toma el valor siguiente:

$$y = 0{,}3228x + 0{,}59$$

Atendamos al significado de a y b en la función de la recta. Como ya escribí, b es naturalmente el valor que tiene y cuando x es cero, es decir, el punto en el que se cortan la línea de tendencia y el eje y (se puede ver perfectamente en el gráfico 5) Por otro lado, a indica la pendiente de la recta. En realidad, es equivalente a <u>la tangente del ángulo</u> que forman la línea de tendencia y el eje x (o cualquiera de sus paralelas, como también puede verse en el gráfico 5), que es nuestro ángulo α.

Recuérdese que en un triángulo rectángulo (y aquí lo forman la línea de tendencia, las líneas horizontales y cualquier línea vertical paralela al eje y), la tangente del ángulo α, que se abrevia «tan α» es la proporción entre el cateto opuesto (al ángulo α) y el cateto contiguo (al ángulo α) de un triángulo rectángulo.

En general, el valor de la tangente de un ángulo está entre $-\infty$ y $+\infty$ (puede tener cualquier valor entre menos infinito y más infinito). Sin embargo, en nuestro caso, como se trata de ángulos del denominado primer cuadrante (porque el rendimiento del cuartil superior es mayor que el del cuartil inferior y las líneas de tendencia son crecientes), el valor está más bien entre 0 y $+\infty$. Podemos concretar todavía más. Aunque en teoría el valor pudiera ser muy elevado ($+\infty$), en realidad, a no superará el valor 1,6 (que es el de la línea de tendencia en el caso extremo de que el porcentaje de no bajo rendimiento en los dos cuartiles inferiores fuera del

0% y en los dos cuartiles superiores del 100%). En el caso concreto de los países que realizaron la prueba PISA en el año 2015, <u>tan α</u> oscila entre 0,0039 (que corresponde a los resultados de Macao) y 0,6804 (que corresponde a los de Perú). Así pues, con los valores de los cuartiles podríamos construir <u>un índice de desigualdad</u>, de carácter comparativo, y aplicarlo a todos los países que se realiza la prueba PISA, y todo ello a partir de la <u>pendiente</u> de la línea de tendencia de los valores de los cuartiles. Simple, pero eficaz.

Si usted ya ha comenzado a practicar el ejercicio de pensar relaciones imaginando rectas y ángulos, tal vez se haya planteado la cuestión siguiente: en el asunto anterior ¿se puede utilizar la obertura del ángulo en lugar de la medida de la tangente? Sí, efectivamente, y el resultado sería <u>el mismo</u>. Una tangente de 0,3228 corresponde a un ángulo de 17º 53' 24", que es lo que se denomina <u>arcotangente</u> (que se abrevia: «arctan»). Lo puede averiguar con una calculadora o, incluso, con páginas web, que permiten el cálculo on line.

Es muy importante poder ofrecer los resultados de la investigación educativa de manera simple, pero eficaz. De nada sirven resultados complicados. Tan simple es el valor de la tangente (tan) como los grados del ángulo (arctan).

Lo importante aquí es el establecimiento de un índice a partir de la pendiente de la línea de tendencia (ya sea como «tan» o como «arctan») que permita <u>comparar</u>, y la comparación es un paso firme hacia la ciencia. Anote también esto: <u>si no podemos comparar, no estamos en el reino de la ciencia</u>.

Fíjese lo que hemos hecho: hemos reducido los cuatro datos de unos cuantiles a un único <u>indicador</u> («tan» o «arctan») y eso nos permite comparar mejor que si vamos arrastrando muchos datos. Más adelante se expondrán otros procedimientos análogos.

Hasta aquí estoy escribiendo sobre líneas «rectas» como si fuera algo sencillo. ¿Que pasaría si le dijera que no existen las líneas rectas, que solo existen líneas curvas?

11.- HABITAMOS EN LAS CURVAS

Pero ¿qué es una curva?

Tal vez usted, cuando haya leído la pregunta «¿qué es una curva?», ha tenido la tentación inmediata de señalar un objeto curvado de su entorno o dibujar un trazo curvo (un arco) en un papel, y puede que haya pensado o haya dicho: «eso es una curva». Es cierto, pero solo <u>en parte</u>. Porque también si ha señalado algún objeto de perfil rectilíneo o ha trazado una línea recta, eso <u>también</u> es una curva. Una línea recta no es más que un tipo de curva (por ejemplo, un arco de circunferencia de radio infinito).

Dos ejes perpendiculares, un plano cartesiano, unos puntos... ¿fácil? Lógicamente, a estas alturas del texto, usted ya tiene buenos motivos para desconfiar de algo que parece tan sencillo (¡hágalo!). Porque, ¿no son también los ejes rectilíneos dos curvas, dos arcos de radio infinito? Y una curva, ¿no es una sucesión de puntos? Y el plano «cartesiano», ¿no será también, en tanto que plano, un tipo especial de superficie curva? Pues sí, así es. Tal vez ahora se entienda mejor la frase que leyó en la introducción: «No hay aquí respuestas sino cuestiones para aprender a indagar».

Pero además, una de las consecuencias de la teoría de la relatividad general formulada por Albert Einstein, es que habitamos en la curvatura del espacio-tiempo. Las curvas no están fuera, sino que vivimos en ellas. Por tanto, en nuestro universo curvo no puede extrañarnos que las relaciones entre dos variables adopten precisamente esta forma.

Volvamos a nuestro ejemplo inicial: ¿Cómo se relacionan las variables altura y edad?

Naturalmente, esta no es una relación lineal: ¡Todos los años no crecemos lo mismo! Por tanto, la respuesta a la pregunta es fácil: si guardan una relación, esta se podrá expresar mediante una curva (adviértase que no se dice que la relación sea la curva, sino que se expresa mediante ella[20]).

Por ejemplo, la imagen 6 es un ejemplo de una curva que relaciona edad y altura. Está elaborada por la Organización Mundial de la Salud y se refiere a niños (no niñas) entre 0 y 2 años.

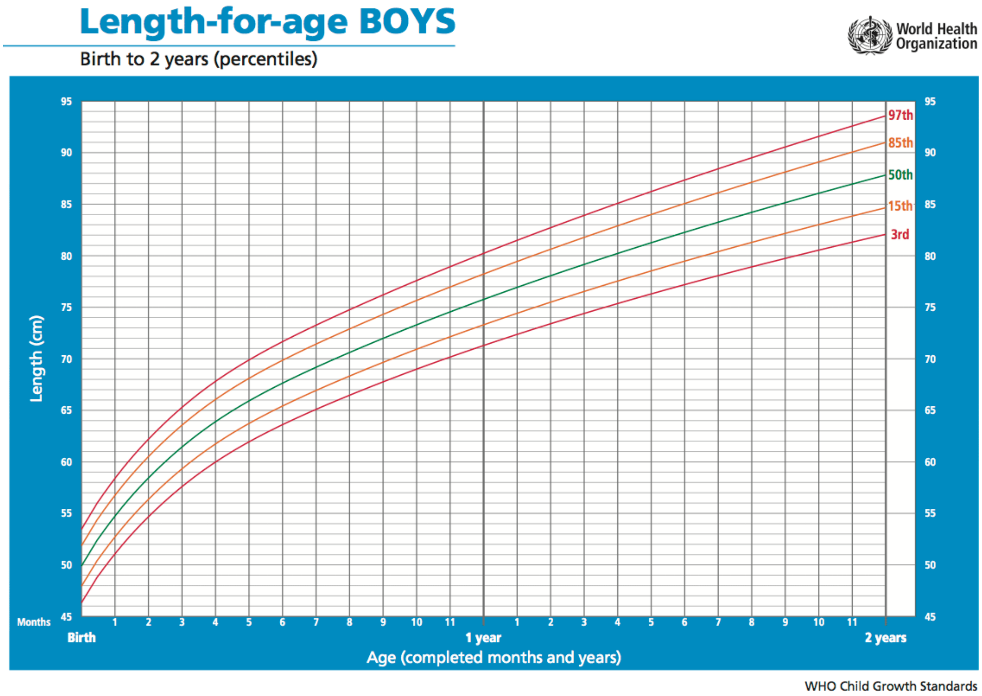
Fíjese además que en este gráfico en realidad se presentan las curvas de los percentiles 3º, 15º, 50º, 85º y 97º.

Pues bien, haré aquí una afirmación tajante: no entraremos en el ámbito de la ciencia si no podemos representar algo en unos ejes cartesianos. Y lo que se representa tampoco es la ciencia misma, sino que lo que se expresa es una relación, la que queda invisible para los ojos, como decía El Principito.

---

[20] Por eso he hablado de lenguaje matemático. Sobre esto debatieron interminablemente los filósofos medievales, con el ejemplo de la palabra «rosa»: ¿Qué relación hay entre la rosa y la palabra «rosa»? Una de los filósofos más importantes clarificador en estos debates fue Guillermo Ockam. Por esta razón, el novelista Umberto Eco llamó Guillermo al protagonista de su novela *El nombre de la rosa*.

[imagen 6]



Fuente: https://www.who.int/childgrowth/standards/h_f_a_tables_z_boys/en/

Naturalmente uno puede indagar la realidad natural o social de muchas maneras y no todas ellas tienen que utilizar necesariamente nociones cuantificables. Además, hay conceptos científicos que no podemos expresar como un número. Por ejemplo, la «vida» es un concepto científico (y muy importante hasta el punto que da nombre a una ciencia: la biología, del griego «*bios*» que significa «vida»), y no podemos decir que alguien tiene «3» o «5» de vida (salvo que estemos jugando a un videojuego, pero esto es otra cosa). Pero estos dos argumentos no invalidan lo dicho antes. Porque, por una parte, se pueda indagar una realidad natural o social con técnicas no cuantitativas (por ejemplo, mediante técnicas cualitativas) e incluso puede ser que se alcance de este modo un repertorio de tesis verdaderas, pero eso no significa que tengamos que considerarlas ciencia, del mismo modo que «observo por la ventana que el sol brilla» puede ser una afirmación verdadera (e, incluso, muy útil o muy inspiradora), pero no por ello forma parte de la ciencia. Y, por otra parte, es cierto que hay conceptos científicos que no podemos traducir a un número[21], pero la biología expresa relaciones que sí que se pueden representar en ejes cartesianos: por ejemplo, la relación entre la cantidad de nutrientes y el crecimiento de una planta.[22]

Por eso, el primer consejo para cualquier investigación sobre la educación es que, más pronto que tarde, intente hacer alguna curva en unos ejes cartesianos, intente expresar de esa manera lo que quiere investigar.

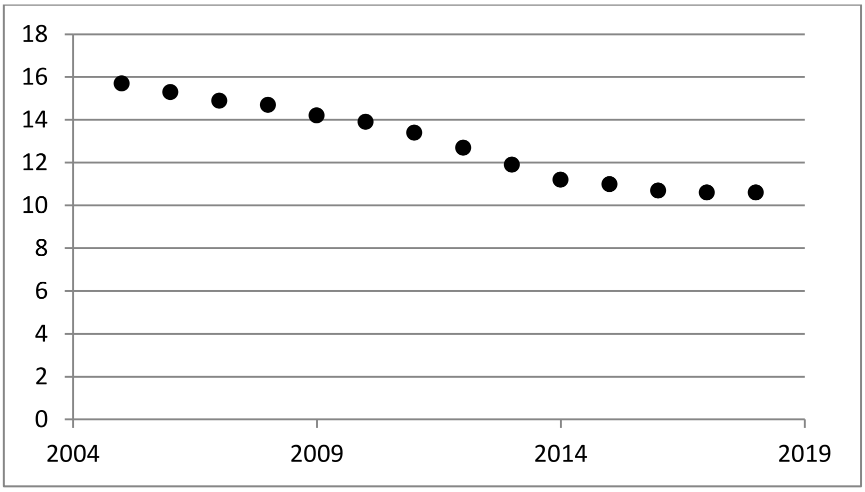
[21] Nicholas Georgescu-Roegen, cuyos textos inspiran este recuadro, habló de conceptos que no eran «aritmomórficos» (es decir, con forma numérica).
[22] Dejo aquí de lado el debate sobre las ciencias hermenéuticas. Sobre esto Usted puede leer mi texto: «Fundamentos filosóficos y sociológicos de la investigación biográfica», que encontrará aquí: <https://www.uv.es/fjhernan/uab.pdf>

12.- TRAZANDO CURVAS

Si comprendió (o ya había comprendido antes de comenzar a leer este texto) que la ciencia expresa relaciones que generalmente podemos representar en ejes cartesianos[23] mediante curvas, podemos dar el siguiente paso de nuestro argumento, que no resulta fácil. Por ello utilizaremos otro ejemplo del ámbito educativo.

El gráfico 6 representa el porcentaje de abandono educativo temprano (es decir, el porcentaje de estudiantado que no cursa educación secundaria postobligatoria) en el conjunto de la Unión Europea entre los años 2005 y 2018. En el eje *x* (el eje de las abscisas o eje horizontal) tenemos los años de la serie y en el eje *y* (el eje de las ordenadas o eje vertical) los porcentajes de abandono educativo temprano del conjunto de la Unión Europea[24].

[Gráfico 6]



Fuente: Eurostat cf. <https://ec.europa.eu/eurostat> código: tesem020 (consulta 24/02/2019)

¡De nuevo tenemos aquí las moscas de Descartes! Si alguien dijera, por ejemplo, que el abandono educativo temprano desciende no estaría haciendo ciencia, porque no estaría expresando ninguna relación[25] (igual que no es ciencia decir que por mi ventana veo brillar el sol o componerle un lindo poema), simplemente estaría comentando una gráfica. Para comenzar a hacer ciencia tengo que, como he propuesto, trazar una <u>curva</u>. ¡¿Una?! En realidad se pueden trazar <u>infinitas</u> curvas que pasen por esos puntos Podría distraerse cogiendo un bolígrafo y trazando una

[23] Con ejes de dos o más dimensiones, claro.

[24] La misma página de Eurostat (que es agencia de estadística de la Unión Europea) permite descargar los datos en diversos formatos.

[25] El filósofo Immanuel Kant hablaba, con razón, de la ciencia como una relación «universal y necesaria».

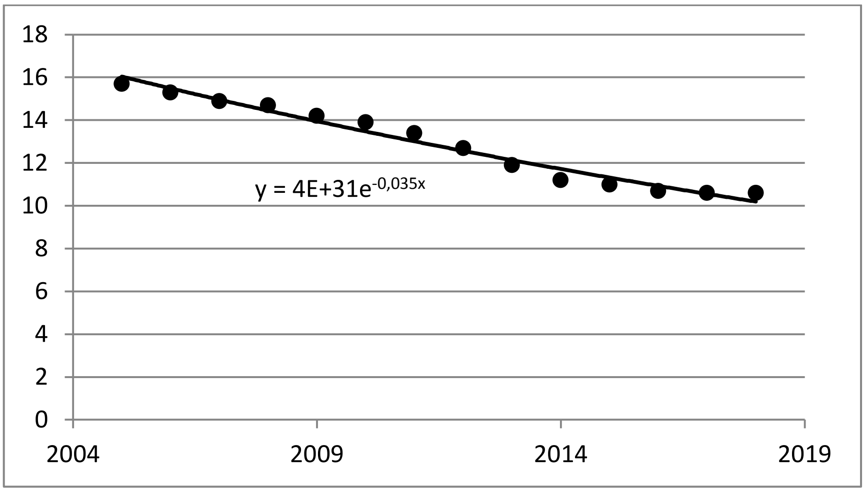
espiral desde un punto central al que estuviera a la derecha y luego al de la izquierda y así sucesivamente hasta unirlos todos. Y luego comenzando por otro punto y siguiendo otro orden. Después de ese ejercicio (relajante, sin duda) habrá llegado a la conclusión de que en ese plano se pueden trazar infinitas curvas que pasen por los puntos marcados (antes ya escribí sobre cómo podemos trazar infinitas rectas). Pues vamos a ir viendo las que se pueden trazar con la hoja de cálculo.

Le advierto que las largas explicaciones que siguen son un poco arduas. Si quiere, puede saltárselas y volver a ellas cuando lo precise.

Línea de tendencia exponencial

Procedemos como hemos antes con la línea de tendencia, pero en lugar de escoger en la ventana de diálogo de la imagen 5 la opción «lineal», comenzaremos con la primera opción: «exponencial». Introduciremos también la ecuación en el gráfico. El resultado es el gráfico 7.

[Gráfico 7. Línea de tendencia exponencial]



Una curva exponencial tiene una función con la forma:

$$y = ae^{bx}$$

Donde a y b son números cualesquiera y e es un número constante (e es un número irracional, que sirve de base para los logaritmos naturales o neperianos que equivale a: 2,718281828459045235360287471352 7... [26] –tiene un número infinito de decimales –).

En el caso anterior, la ecuación de la línea de tendencia es:

$$y = 4E{+}31e^{-0,035x}$$

Esta notación resulta un poco inhabitual, por lo que hay que explicarla.

4E+31 significa: 4 por 10 elevado a 31 (es decir un 4 y 31 ceros)

$e^{-0,035x}$ significa el número *e* elevado a -0,035*x*

Vamos a hacer, por ejemplo, la <u>estimación</u> con esta fórmula para el año 2020 (es decir: ¿qué abandono habrá en ese año?). Primero explicaré el procedimiento con una calculadora (física o digital) y luego con una hoja de cálculo.

Procederemos así con la calculadora.

Primero calcularemos el producto -0,035 por *x* (aquí: 2020), tecleando lo siguiente:

[-][0][,][0][3][5] [*] [2][0][2][0] [=]

-70,7

Ahora, elevaremos el número *e* a ese valor. Para ello presionaremos la tecla [Inv] de funciones inversas y después utilizaremos [e^x]. Procederemos así:

[Inv] [e^x]

1,974149914813674101594720725051e-31

Adviértase que en esta fórmula la letra *e* significa «por 10 elevado a» (no el número *e*)[27]. Guardamos ese valor en la memoria

[M+]

Después elevamos 10 a 31 y lo multiplicamos por 4 así:

[1][0] [x^y] [3][1] [=] [*] [4] [=]

40000000000000000000000000000000

(Advertencia: este paso también se puede hacer de otro modo: [4] [Exp] [3][1] [=], pero seguiremos con el otro procedimiento para no complicar el asunto)

[*] [MR] [=]

7,8965996592546964063788829002042

[26] En algunas calculadoras hay una tecla con el número [e]. En otras, se puede conseguir anotando en la pantalla [1] y presionando la tecla [e^x]. A veces esta aparece con [Inv][ln].

[27] Seguro que alguna vez se habrá preguntado: ¿por qué hacen las cosas tan complicadas y unas veces unas letras significan una cosa y otras veces otra? Sí, yo también me lo he preguntado y no tengo respuesta.

Por tanto, el abandono previsto para el año 2020 es de 7,89% (en el gráfico 5 los valores del *y* son porcentajes, aunque no se diga allí).

Si se prefiere utilizar el programa Excel u otra hoja de cálculo, entonces podemos:

Anotar en una celda, por ejemplo A1, el año del que queremos hacer una estimación.

Seleccionar la celda donde queremos obtener el resultado y teclear en la barra de fórmulas (que también expresaré mediante un recuadro):

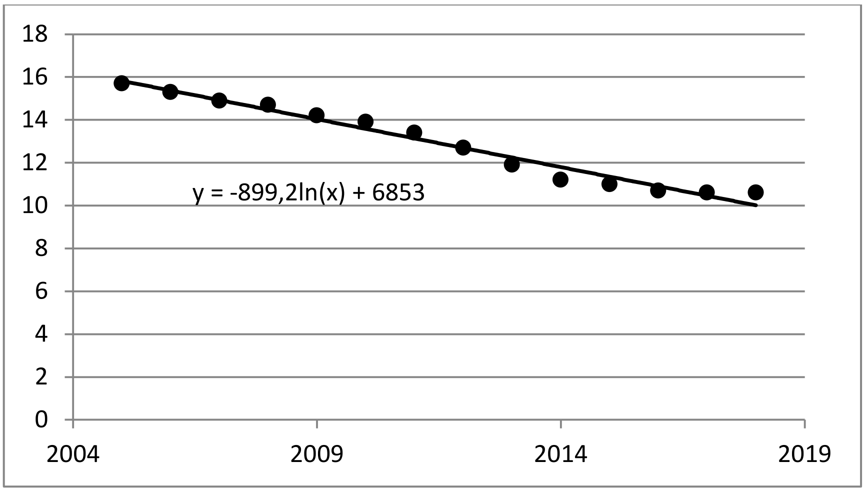
=4*POTENCIA(10;31)*POTENCIA(EXP(1);(-0,035*A1))

(si el año se anota en otra celda, tendrá que sustituir su nombre por el de A1 en la fórmula anterior). Darle a la tecla [ENTER] y listo.

Línea de tendencia logarítmica

Con esta opción se genera el gráfico 7 y la ecuación correspondiente.

[Gráfico 7. Línea de tendencia logarítmica]



Una curva logarítmica tiene una función con la forma:

$$y = a\ ln(x) + b$$

Donde a y b son números cualesquiera y ln es un logaritmo natural o neperiano[28] (un logaritmo que tiene el número e como base). En ese caso la ecuación resultante es:

$$y = -899{,}2 \ln(x) + 6853$$

Vamos a realizar la misma estimación para el año 2020, siguiendo los pasos siguientes con la calculadora:

[2][0][2][0] [ln]

7,6108527903952504443194218430555

[*] [(][-][8][9][9][,][2][)] [=]

-6843,6788291234091995320241212755

(otra posibilidad, para evitar los paréntesis, es multiplicar por 899,2 y después invertir el signo con la tecla [±])

Por último, añadimos el segundo sumando del miembro de la derecha de la ecuación:

[+] [6][8][5][3] [=]

9,3211708765908004679758787244675

Por tanto, el abandono previsto para el año 2020 es de <u>9,32%</u>

Si quiere utilizar una hoja de cálculo, seleccione una celda (por ejemplo, A1) y luego seleccione otra, donde quiera obtener el resultado, y escriba en la barra de fórmulas:

```
=-899,2*LN(A1)+6853
```

Y presiona la tecla [ENTER].

Adviértase que el valor resultante es ligeramente <u>diferente</u> al anterior ¡porque en definitiva estamos con dos modelos diferentes –la curva exponencial y la curva

---

[28] Recuerde que un logaritmo es, por así decir, lo inverso de una potencia. Si, por ejemplo, 10^3=1000, entonces el logaritmo en base 10 de 1000 es 3, lo que se expresa así: $\log_{10} 1000=3$. Los logaritmos naturales o neperianos tienen como base el número e. Por ejemplo: $\log_e 1000=6{,}9077...$, porque $e^{6,9077...}$ o, si prefiere, $2{,}7182...^{6,9077...}$, es igual a 1000, aunque no se escribe log_e sino ln.

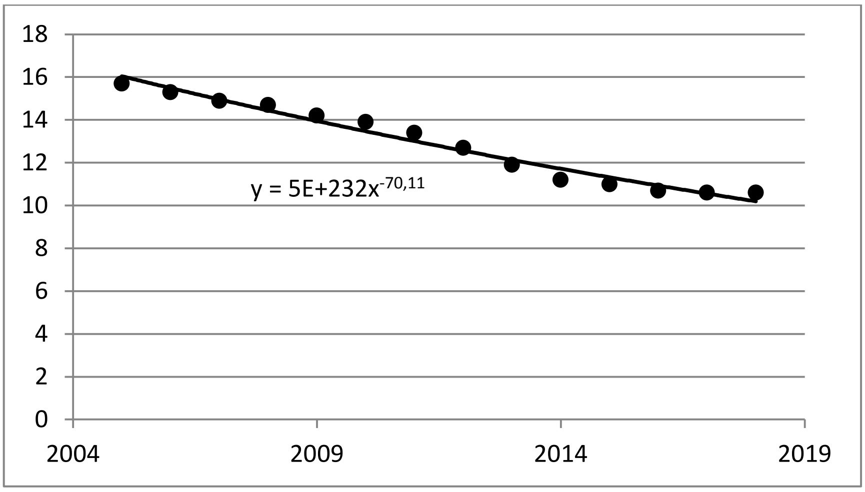
logarítmica–. Si continuamos con las otras curvas obtendremos otros resultados distintos. Si esto le incomoda, vuelva a leer el epígrafe 6.

Dejaremos ahora de lado la curva polinómica, que explicaré más adelante.

Línea de tendencia potencial

Con esta opción se genera el gráfico 8 y la ecuación correspondiente.

[Gráfico 8. Línea de tendencia potencial]



Una curva potencial tiene una función con la forma:

$$y = ax^b$$

Donde a y b son números cualesquiera. Aquí la función de la curva es:

$$y = 5E{+}232x^{-70,11}$$

Recuerde que esta «E» mayúscula significa aquí «por 10 elevado a».

Realizaremos con una calculadora la estimación para 2020.

[2][0][2][0] [x^y] [(][-][7][0][,][1][1][)] [=]

1,8273237047649572577741553738488e-232

[M+]

También en esta fórmula la letra e significa «por 10 elevado a» (no el número e). Luego calculamos el término a y lo multiplicamos por el registro de la memoria.

[1][0] [x^y] [2][3][2] [*] [5]

5,e+232

(También se puede hacer este paso con la tecla [Exp], como se explicó antes)

[*] [MR]

9,136618523824786288870776869244

Por tanto, el abandono previsto según esta curva para el año 2020 es de 9,13%.

Para hacer la estimación con una hoja de cálculo, como Excel, hay que seleccionar una celda (por ejemplo: A1), en la que escribiremos el año, y en la que queramos obtener el resultado, seleccionarla y escribir en la barra de fórmulas:

=-899,2*LN(A1)+6853

Línea de tendencia polinómica

Antes de mostrar el gráfico, hay que recordar que un polinomio puede presentar muchos (en realidad, infinitos) grados (el programa Excel habla de *orden*).

Grado 0 $y = a$

Grado 1 $y = ax+b$

Grado 2 $y = ax^2+bx+c$

Grado 3 $y = ax^3+bx^2+cx+d$

Grado 4 $y = ax^4+bx^3+cx^2+dx+e$

Grado 5 $y = ax^5+bx^4+cx^3+dx^2+ex+f$

Grado 6 $y = ax^6+bx^5+cx^4+dx^3+ex^2+fx+g$

Etc.

Téngase en cuenta que, en todos estos casos, el último sumando es el producto de un número por x^0, porque $x^0 = 1$. Y el penúltimo sumando es un número multiplicado por x^1, porque $x^1 = x$. Por eso, algunas personas prefieren escribir en orden inverso los sumandos del miembro de la derecha o utilizar una misma letra con subíndices (que coinciden entonces con las potencias de x). Así, por ejemplo, el polinomio de grado 6 sería:

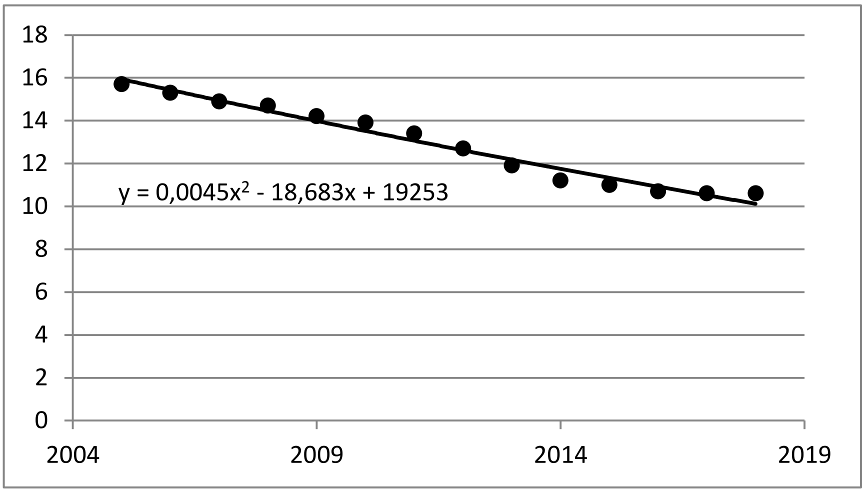
$$y = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4 + a_5x^5 + a_6x^6$$

Esta es otra manera de «notación», pero el significado es el mismo. Aquí usaremos la expresión inicial (la que presenta las potencias decrecientes).

Veamos el caso de la ecuación polinómica de orden (grado) 2.

(La ecuación polinómica de grado 1 sería la ecuación de la recta, que ya ha sido comentada).

[Gráfico 9. Línea de tendencia polinómica. Grado 2]



Realizaremos la estimación para el año 2020 y obtenemos un valor de -124,86% que claramente se desvía de las estimaciones anteriores (e incluso no resulta adecuado, por cuanto no hay abandono «negativo»). Con una hoja de cálculo, utilizaríamos la fórmula:

```
=0,0045*A1*A1-18,683*A1+19253
```

Que nos ofrece el mismo resultado divergente. Por ello, tenemos que tomar precauciones con las estimaciones a partir de curvas polinómicas.

Tal vez algunas personas se inquieten con ecuaciones tan largas como las que hemos visto. Hay que recordar que con las fórmulas matemáticas sucede lo mismo que con las serpientes: es preciso tratarlas con cuidado y las más largas no siempre son las peores. Así que no nos preocupemos de la apariencia de las fórmulas (¡son inofensivas!).

Ha de quedar claro que, si tenemos una función, podemos hacer una previsión (como acabamos de ver). Ahora podemos reconstruir el camino y entender la importancia de la curva: ella, o mejor aún, su función, es el puente que nos ha permitido pasar de los datos hasta la previsión.

13.- CURVAS ELEGANTES

Veremos a continuación un ejemplo de cómo podemos usar las líneas de tendencia para obtener buenos resultados. En una investigación de Alícia Villar, Rafael García y el autor de este texto sobre trayectorias universitarias, en la que analizamos las bases de datos de la Universitat de València (con unos 50.000 estudiantes cada curso) desde el curso 2009/10 hasta el curso 2014/15[29], localizamos los estudiantes que cambiaron de grado (se reubicaron en otra carrera) y realizamos un recuento de los que lo habían hecho en el primer año a partir de su primera matrícula, en el segundo, etc. Los resultados se muestran en la tabla 4.

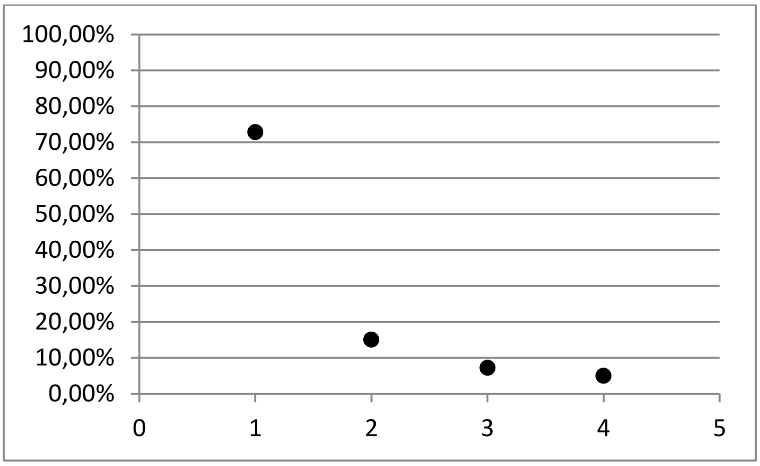
Tabla 4.- Estudiantes que cambiaron de grado. Universitat de València 2009/2010-2014/2015. Valores absolutos y porcentajes

	Estudiantes que cambiaron de grado	Porcentajes
En el año 1º	1789	72,75%
En el año 2º	371	15,09%
En el año 3º	177	7,20%
En el año 4º	122	4,96%
Total	2459	100,00%

Fuente: Elaboración propia con datos de la Universitat de València.

Podemos cargar estos datos en una hoja de cálculo y pedirle al programa que inserte un gráfico de dispersión. Esto es lo que se muestra en el gráfico 10.

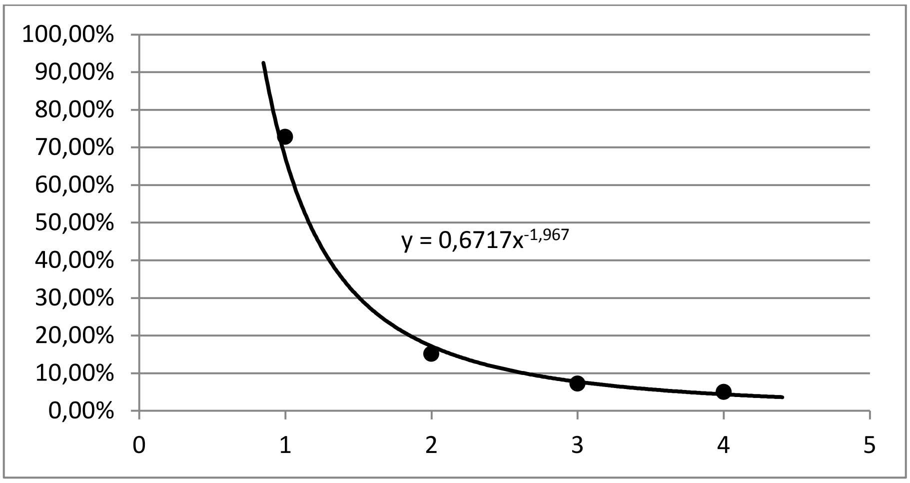
[Gráfico 10]



[29] Recuérdese que en el hemisferio Norte, los cursos académicos se mencionan con dos años naturales.

Siguiendo el procedimiento explicado, podemos trazar la línea de tendencia que más se aproxime a los puntos, con el fácil recurso de ir probando las diversas opciones que nos proporciona el programa: exponencial, logarítmica, polinómica, de diversos grados, o potencial (recuérdese la ventana de diálogo de la imagen 5). En este caso he elegido la función potencial. De este modo, se traza la línea de tendencia que aparece en el gráfico 11 y se le pide al programa que inserte la ecuación que corresponde a esa línea de tendencia.

[Gráfico 11]



Recuérdese que una función potencial es una ecuación del tipo:

$$y = ax^b$$

En concreto, la línea trazada en el gráfico 11 presenta la ecuación:

$$y = 0{,}6717x^{-1{,}967}$$

Tenemos aquí otra ecuación, breve, pero no por ello sencilla (una serpiente pequeña, como decíamos antes), que relaciona valores de x e y. ¿Qué podemos hacer con esta ecuación? Podemos intentar darle una forma más elegante.

Para pensar matemáticamente es muy importante intentar que los resultados sean elegantes. De hecho, fueron los discípulos de Pitágoras, en la antigua Grecia, los que defendieron, tal vez por primera vez, que las formulaciones matemáticas tenían que presentar esa elegancia que las asociaba con la armonía musical. De hecho, si usted contempla el gráfico 10 verá que guarda una cierta semejanza con una partitura musical donde alguién hubiera anotado cuatro notas negras.

En el caso de la ecuación anterior, darle una forma más elegante es relativamente fácil. Por un lado, 0,6717 es un número muy próximo a 0,6666..., que equivale a $\frac{2}{3}$.
Por otro lado, −1,967 es un número muy cercano a −2. Por ello, podemos reescribir la fórmula anterior como la siguiente:

$$y = \frac{2}{3}x^{-2}$$

Esta fórmula es más elegante, se puede recordar más fácilmente, queda mejor en un trabajo fin de grado o de máster, en una tesis doctoral o en un artículo científico y, además, el margen de error con la fórmula anterior es despreciable en el cálculo.

En la tabla siguiente, se puede apreciar que los errores de utilizar la ecuación de la línea de tendencia o su conversión en una forma más elegantes són muy bajos, inferiores al 0,5%.

Año	Función original	Función "elegante"	Error
1	0,6717	0,6667	0,50%
2	0,1718	0,1667	0,51%
3	0,0774	0,0741	0,33%
4	0,0439	0,0417	0,23%

Veamos ahora otro ejemplo de utilización de las líneas de tendencia en la investigación. En la tabla 5 hay una serie de países europeos y los datos de dos variables: la inversión pública en educación (como porcentaje de los presupuestos públicos en educación respecto del producto nacional bruto[30]) y el porcentaje de estudiantado de 15 años con bajo rendimiento en matemáticas (según las pruebas PISA). En ambos casos, los datos están referidos al año 2012.

Tabla 5.- Inversión pública en educación y estudiantado con bajo rendimiento en matemáticas. 2012.

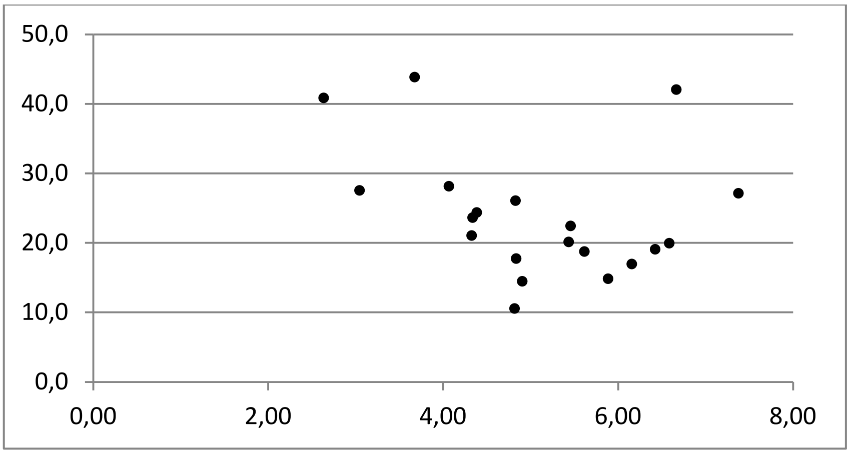
País	Inversión pública	Bajo rendimiento
Alemania	4,84	17,7
Austria	5,62	18,7
Bélgica	6,43	19,0
Bulgaria	3,68	43,8
Chequia	4,33	21,0
Chipre	6,67	42,0
Eslovaquia	3,05	27,5
Eslovenia	5,44	20,1
España	4,34	23,6
Estonia	4,82	10,5
Francia	5,46	22,4
Holanda	5,89	14,8
Hungría	4,07	28,1
Irlanda	6,16	16,9
Letonia	6,59	19,9
Lituania	4,83	26,0
Luxemburgo	4,39	24,3
Polonia	4,91	14,4
Rumanía	2,64	40,8
Suecia	7,38	27,1

Fuente: Eurostat, códigos: educ_uoe_fine06 y sdg_04_40.

Trasladamos esta tabla 5 a una hoja de cálculo e insertaremos un gráfico. Obtenemos así el gráfico 12.

[30] Las inversiones o gastos en educación de un país se suelen contabilizar en términos absolutos (el montante económico) o en términos relativos, bien la inversión por estudiante o bien la inversión respecto del conjunto de bienes y servicios producidos en un año, es decir, el producto interior o nacional bruto.

[Gráfico 12]



Como hemos explicado, para trazar una línea de tendencia ponemos el cursor sobre uno de los puntos del gráfico y hacemos clic en el botón derecho del ratón. En el menú que se despliega elegimos el tipo de línea de tendencia y le solicitamos al programa que también nos presente la ecuación de la línea (imágenes 4 y 5). Los gráficos que podemos obtener y las ecuaciones se muestran a continuación.

[Gráficos 13-21]

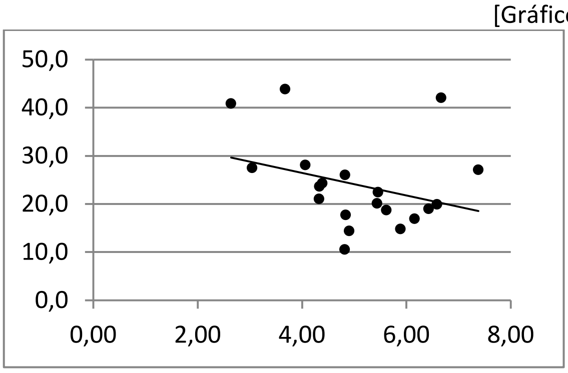


Gráfico 8.- Línea de tendencia lineal (recta)

$y = -2{,}3539x + 35{,}881$

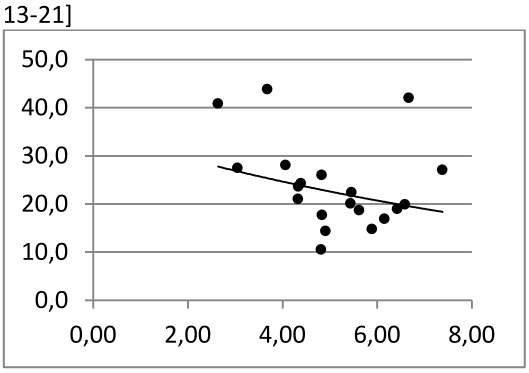


Gráfico 9.- Línea de tendencia exponencial

$y = 34{,}931e^{-0{,}087x}$

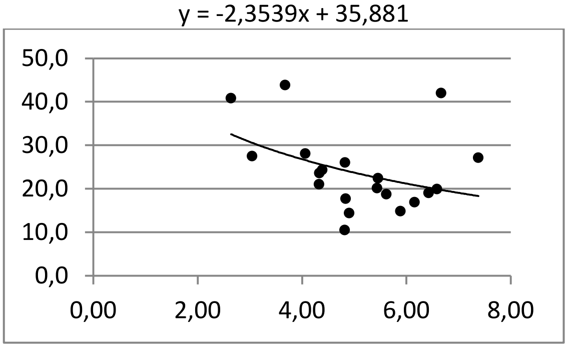


Gráfico 10.- Línea de tendencia logarítmica

$y = -13{,}82\ln(x) + 45{,}95$

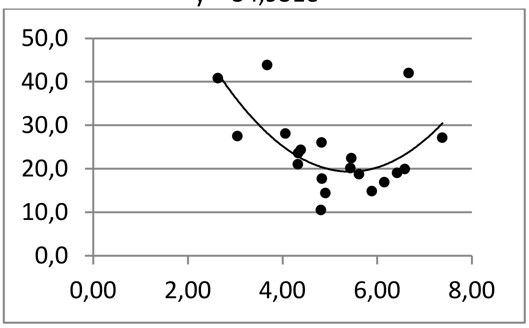
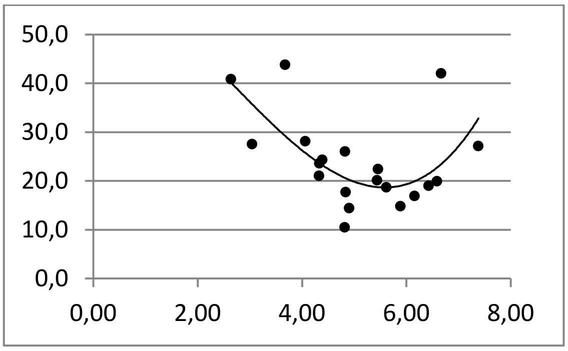


Gráfico 11.- Línea de tendencia polinómica (orden 2)

y = 2,9098x2 - 31,554x + 104,88



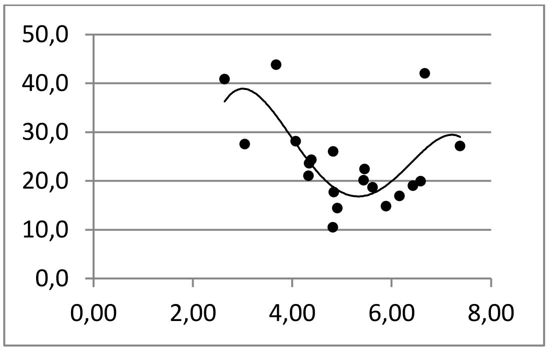
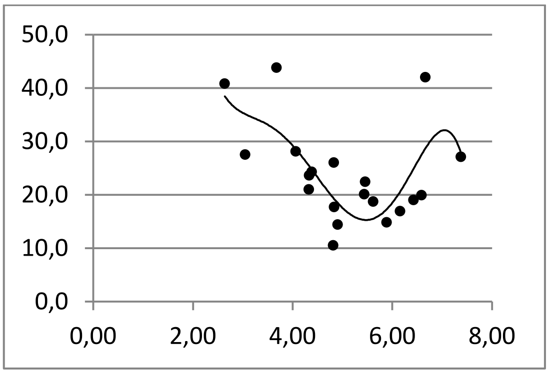


Gráfico 12.- Línea de tendencia polinómica (orden 3)

y = 0,4054x3 - 3,1267x2 - 3,021x + 62,436

Gráfico 13.- Línea de tendencia polinómica (orden 4)

y = -0,8665x4 + 17,958x3 - 131,87x2 + 400,09x - 389,2



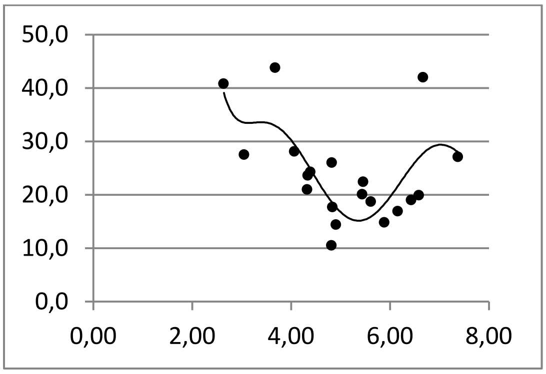


Gráfico 14.- Línea de tendencia polinómica (orden 5)

$y = -0,5161x^5 + 12,141x^4 - 109,71x^3 + 476,65x^2 - 1005,1x + 864,92$

Gráfico 15.- Línea de tendencia polinómica (orden 6)

$y = 0,2317x^6 - 7,4373x^5 + 96,39x^4 - 643,96x^3 + 2335,6x^2 - 4366x + 3328,9$

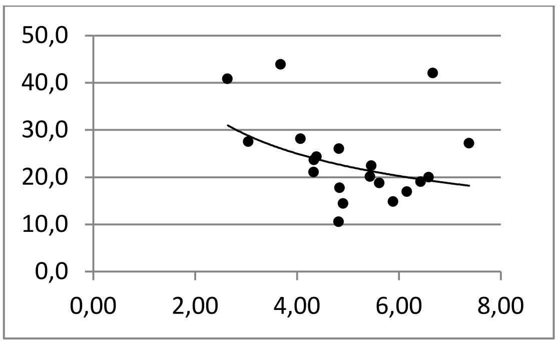


Gráfico 16.- Línea de tendencia potencial

$y = 51,186x^{-0,517}$

¿Que función deberíamos elegir? Esta cuestión, como hemos dicho, no se puede resolver automáticamente. Antes de profundizar en esto tenemos que tomar una precaución: preguntarnos si todos los valores son correctos o hay valores divergentes que estén alterando la tendencia. Buscaremos si hay «malos bailarines»

15.- EMPIEZA EL BAILE

Me gustaría que usted observara esas gráficas como si estuviera contemplando una pista de baile desde arriba. Los puntos serían los bailarines (o a las moscas de Descartes, revoloteando sobre la mesa). Podemos imaginar también que las diferentes líneas de tendencia son diversas coreografías: ¿qué bailan nuestros bailarines? ¿samba, forró, tango, cumbia, bachata?

Un procedimiento más sofisticado que la línea de tendencia para establecer la relación entre dos variables es calcular su coeficiente de correlación, aquí utilizaremos el coeficiente de Pearson-Bravais (para simplificar, de Pearson).

El coeficiente de correlación o coeficiente de Pearson se suele expresar mediante la letra r , la mayúscula, R, o la letra griega equivalente ro, ρ; aquí utilizaremos R, que es como aparece en el programa Excel.

El coeficiente de Pearson expresa la relación entre dos variables, x e y. El valor de R puede oscilar entre -1 y 1.

En el caso de que haya una correlación intensa y directa entre x e y, el valor de R se aproxima a 1. Esto quiere decir que cuando los valores de una variable crecen, los correspondientes de la otra suelen crecer al mismo ritmo, y viceversa.

Por otra parte, en el caso de que no exista prácticamente correlación, el valor de R se aproxima a 0. Entonces, la oscilación de una variable no se corresponde con la de la otra.

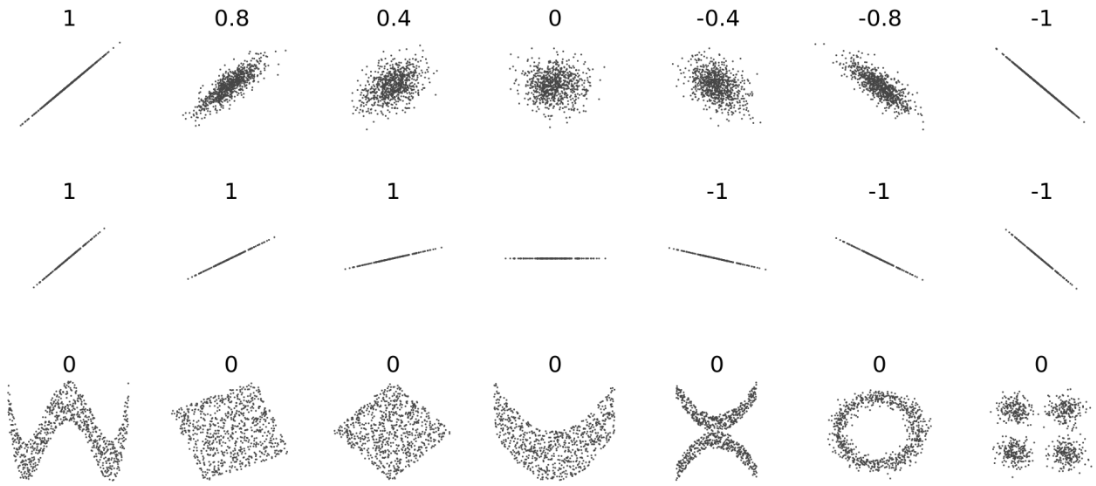
Por último, en caso de que una variable aumente y la otra disminuya al mismo ritmo, el valor de R se acerca a -1. Hablaremos entonces de una correlación intensa, pero inversa. Sin embargo, simplemente con emplear la variable inversa de una de las dos (digamos x^{-1} o y^{-1}), R cambia su signo. Por ello, muchas veces podemos utilizar el valor absoluto de R (lo que se escribe: |R|), sin tener en cuenta si es positivo o negativo su valor. Por tanto:[31]

$$-1 \leq R \leq 1 \qquad ; \qquad 0 \leq |R| \leq 1$$

[31] Recuerde que el signo « ≤ » significa: menor o igual que.

La Wikipedia presenta un ilustración que resulta interesante, en la que se relacionan las nubes de puntos de un gráfico XY (imagínese los ejes cartesianos, porque no están en la ilustración) con los valores de R. Lo malo, como pasa siempre, es que la vida no es tan simple como se nos muestra en los diccionarios y las nubes de puntos pueden adoptar más formas. Compare, por ejemplo, los gráficos 12-21 con los de la ilustración (imagen 7).

[Imagen 7]



(El valor de R aproximado está sobre el dibujo de la nube de puntos)

Una definición más rigurosa del Coeficiente de Correlación de Pearson sería la siguiente:

El coeficiente de correlación de Pearson se define como el cociente de la *covarianza* de dos variables, *X* e *Y*, por el producto de sus *desviaciones típicas*.

$R_{xy} = \frac{\sigma_{xy}}{\sigma_x \, \sigma_y}$

Veamos estos conceptos.

La *covarianza*, σ_{xy}, se define como el sumatorio del producto de la diferencia de cada valor de *x* menos su media por el correspondiente valor de *y* menos su media, dividido por el número de valores -1.

$\sigma_{xy} = \frac{\Sigma \, (x_i - \bar{x})(y_i - \bar{y})}{n-1}$

La *desviación típica*, que se expresa con la letra *s* o la letra griega equivalente *sigma*, σ, equivale a la raíz cuadrada de la *varianza*. La varianza es la suma de la diferencia de cada valor con la media dividida por el número de valores.

$\sigma_x^2 = \frac{1}{n} \Sigma \, (x_i - \bar{x})$

El coeficiente de correlación, R, se puede calcular manualmente o mecánicamente. Los cálculos manuales son muy largos, pero afortunadamente nuestros amigos inhumanos, los ordenadores, permiten calcularlo de un modo sencillo y rápido.

Una de las fórmulas más sencillas para el cálculo manual es la siguiente;

$$R_{xy} = \frac{n\, \Sigma\, x_i\, y_i - \Sigma\, x_i\, \Sigma\, y_i}{\sqrt{n\, \Sigma\, x_i^2 - (\Sigma\, x_i)^2}\ \sqrt{n\, \Sigma\, y_i^2 - (\Sigma\, y_i)^2}}$$

Dados *n* valores de las variables X e Y, para calcular manualmente un coeficiente de correlación de este modo, hay que averiguar los datos sombreados en el esquema.

X	Y	XY	X^2	Y^2
x_1	y_1	x_1y_1	x_1^2	y_1^2
x_2	y_2	x_2y_2	x_2^2	y_2^2
…	…	…	…	…
x_n	y_n	x_ny_n	x_n^2	y_n^2
$\Sigma\, x_n$	$\Sigma\, y_n$	$\Sigma\, x_ny_n$	$\Sigma\, x_n^2$	$\Sigma\, y_n^2$
$(\Sigma\, x_n)^2$	$(\Sigma\, y_n)^2$			

Pero no se esfuerce en calcularlo manualmente. Es muy largo y lo digo por experiencia. La maquina lo hace instantáneamente.

Con una hoja de cálculo hay dos procedimientos para calcular un coeficiente de correlación de Pearson (R). Veamos los dos.

Primer procedimiento

El primer procedimiento consiste en insertar un gráfico (tal como se ha explicado anteriormente), y en el cuadro de diálogo sobre la línea de tendencia hacer clic en <Presentar el valor R cuadrado en el gráfico> (véase la imagen 5, en la parte inferior).

Naturalmente, si disponemos del valor de R^2, podemos obtener R haciendo la raíz cuadrada con una calculadora. Pero atención porque aquí puede producirse un error.

Recuerde usted que, por ejemplo, la raíz cuadrada de 9 es 3, pero también -3 (porque $-3 \times -3 = 9$). Por ello, si disponemos de la raíz de R^2 tenemos que reflexionar sobre si R tendrá signo positivo o negativo. El criterio es:

$\sqrt{R^2}$ tiene signo positivo si la línea de tendencia es creciente.

$\sqrt{R^2}$ tiene signo negativo si la línea de tendencia es decreciente.

¿Y cómo sabemos si la línea de tendencia es creciente o decreciente? Lógicamente, mirando la representación de la línea de tendencia. Si el lado izquierdo tiene valores superiores que el derecho, entonces es creciente, y viceversa. La explicación es sencilla. Recuerdese que la ecuación de la recta tiene la forma:

$$y = ax + b$$

Por ello, si a presenta un valor positivo, la función es creciente y $\sqrt{R^2}$ tiene signo positivo. Si a presenta un valor negativo, la función es decreciente y $\sqrt{R^2}$ tiene signo negativo.

Puede ver como ejemplo los anteriores gráficos 3 o 5, con el factor a positivo y líneas crecientes, o el gráfico 13, con el factor a negativo y la línea claramete decreciente.

Esto resulta bastante sencillo de entender. Si la función es creciente, el factor *a* es positivo y, por lo tanto, cuando crece *x* entonces aumenta el producto *ax* y, por ello, la suma *ax+b*. Pero si *a* es negativo, entonces el producto *ax* también lo es y, sea cual sea el valor de *b*, entonces la suma *ax+b* disminuye a medida que aumenta *x*, por ello la línea tiene carácter decreciente (tiene valores superiores a la izquierda que a la derecha).

Por otro lado, algunos manuales denominan a R^2: «coeficiente de determinación», y entienden que es la proporción de una variable determinada por la otra.

Segundo procedimiento

El segundo procedimiento no exige realizar el gráfico, sino simplemente utilizar las funciones del programa. También aquí hay dos posibilidades:

a) Seleccionar una celda y hacer clic en el botón de funciones [*fx*], que está junto a la barra de fórmulas y luego buscar la correspondiente Al coeficiente de correlación, que responde a la abreviatura:

COEF.CORREL

Aparece entonces una ventana de diálogo. Tenemos que hacer clic en la «matriz» primera y seleccionar con el cursor todos los datos de una variable y luego en la otra «matriz» y seleccionar los de la otra. Haremos clic en [ENTER] y obtendremos el valor de R.

b)La segunda posibilidad es seleccionar una celda y escribir una fórmula directamente en la barra de fórmulas. Si, por <u>ejemplo</u>, los valores de la primera variable están en las celdas B2, B3, B4... hasta B21, y los valores de la segunda variable están en C2, C3, C4... hasta C21, entonces la fórmula será:

```
=COEF.CORREL(B2:B21;C2:C21)
```

Las dos variables tienen que tener el mismo número de datos, lo que no significa que estén en las mismas filas de la hoja de cálculo (como en el ejemplo anterior).

En el caso de la tabla 5, el coeficiente de correlación es R = –0,3191.

En este ejemplo, podríamos decir que R = –0,319 resulta un valor interesante: apunta una cierta correlación (hay correlación, aunque negativa).

¿Resulta suficiente o estará influido el coeficiente por algún valor discrepante (que lo puede ser por razones espúreas)? Intentaremos entonces depurar los datos, esto es... localizar y eliminar a los «malos bailarines».

¿Cómo hacer esto? Los «malos bailarines» son aquellos que se alejan de la línea de tendencia. Por ello, podemos pensar en una línea de tendencia sencilla (por ejemplo (polinómica de orden 2 o potencial) y localizar los puntos más alejados, después eliminaremos esos datos de la hoja y recalculamos R para ver qué pasa.

Pues bien, si eliminamos los valores de Chipre, R alcanza un valor de -0,545. Realmente las estadísticas educativas de Chipre resultan muy discrepantes. Piénsese que Chipre es una república con una parte ocupada militarmente por Turquia (donde proliferan universidades extranjeras, que así tienen una sede en el Espacio Europeo de Educación Superior) y que incluso su capital, Nicosia, está dividida en dos zonas, con tropas de interposición de las Naciones Unidas. Esta subida de R si dejamos de lado Chipre apunta a que nos encontramos ante una correlación importante. Incluso si elimináramos algún otro país, como Suecia o Estonia (que pueden ser divergentes por razones distintas que Chipre), alcanzaríamos una correlación de -0,743 ¡lo que es mucho para una serie de 17 países!

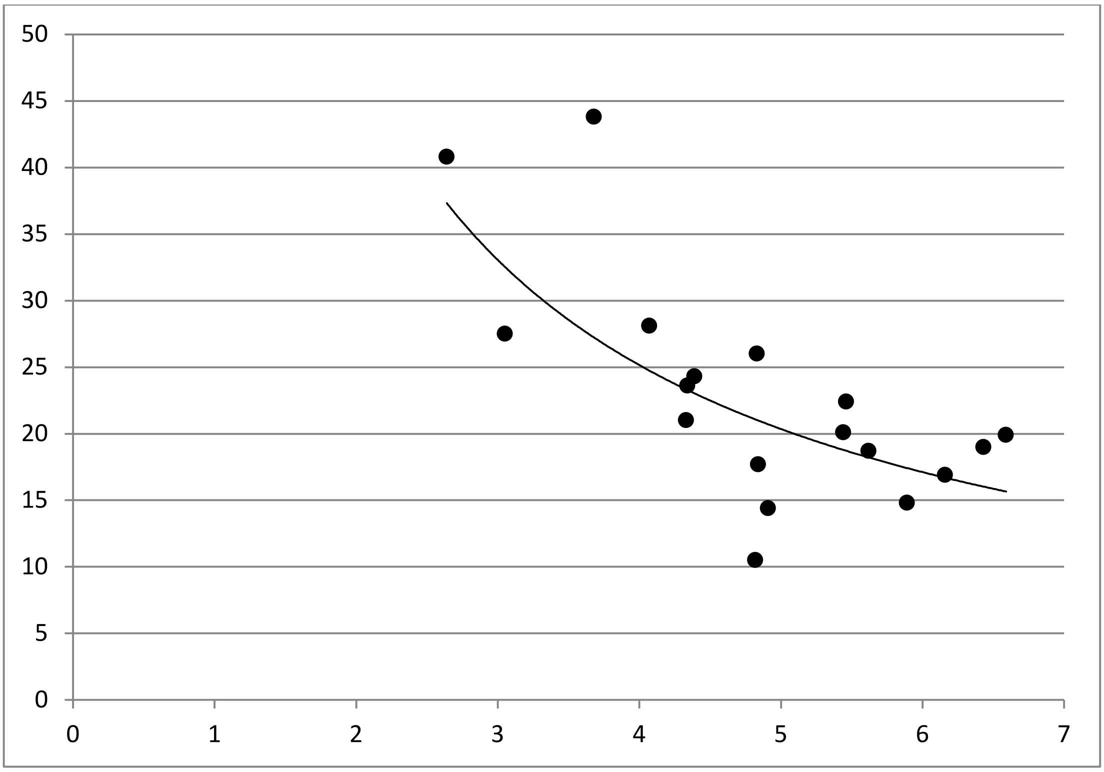
Algunos manuales de estadística proporcionan tablas para poder establecer si una correlación es alta, baja, regular, etc. Pero esto es bastante arbitrario y varía de manual a manual. Además hay que considerar el objeto mismo que estamos investigando.

En general, olvídese de alcanzar valores muy elevados de R en la investigación educativa (por ejemplo, 0,90 o 0,95). Un valor de R algo elevado (por ejemplo, 0,6 o 0,7, como en el ejemplo anterior) ya merece ser celebrado.

Y una recomendación necesaria: si depura los datos, explique las razones por las que lo hace en la memoria de su investigación (como en el caso mencionado de Chipre).

Con los 17 países resultantes podemos volver a trazar una línea de tendencia. Obsérvese como los puntos se aproximan ahora a una línea de tendencia potencial (gráfico 22):

[Gráfico 22]



Ya estamos en condiciones de comenzar a establecer un modelo y transitar por el camino de la ciencia (o si se prefiere: ya entendemos qué están bailando nuestros bailarines, ya intuimos con claridad cuál es la música).

Una línea de tendencia potencial (como su inversa: la logarítmica) sirven para representar fenómenos de saturación. Estos acaecen cuando el incremento (subidas o bajadas) de una variable produce incrementos en la otra, pero cada vez a un menor ritmo.

El ejemplo clásico es el resultado de añadir azucar al café. La primera cucharada supone un incremento de dulzor en la bebida; la segunda, un incremento menor, y así sucesivamente; algunas cucharadas después ya no notaremos que la bebida esté más dulce por más azucar que añadamos[32]. Pues eso mismo parece representar el gráfico 22. A medida que aumenta la inversión pública en educación desciende el bajo rendimiento, pero poco a poco la dinámica se satura, ya no desciende tan rápidamente.[33]

[32] Se trata de un ejemplo incompleto, porque para hacer una gráfica precisaríamos una escala de dulzura.

[33] Hay muchas dinámicas de saturación. Por ejemplo, el efecto de los narcóticos o de algunos medicamentos.

¿Es la inversión pública causa del nivel de rendimiento? Recuérdese que correlación (que aquí es muy elevada) no es causalidad. La causalidad es una palabra «mayor» que tenemos que usar con mucha cautela.

Para poder afirmar la causalidad debemos hacer más operaciones: la primera y más importante reflexionar sobre el vínculo entre un fenómeno y otro. Pensemos: un incremento en la inversión pública determina mejores centros educativos y docentes mejor pagados; se promoverá así una representación general en la que la educación es vista como algo importante para la sociedad, y ello favorecerá que los estudiantes tengan mejores rendimientos, tanto porque gozan de mejores recursos, como porque pueden albergar mayores expectativas. Por lo tanto, parece razonable que la inversión condicione, al menos parcialmente, el rendimiento.

A continuación podemos formular un teorema. Un teorema es una proposición científica que todavía no está suficientemente demostrada. Pero ya hemos avanzado en el camino de la ciencia. Partiremos de la última ecuación de la línea de tendencia:

$$y = 93{,}766\, x^{-0{,}95}$$

Podemos hacer una tabla con valores de x entre 1 y 10 (no hay países con inversión pública por debajo del 1% ni por encima del 10%). Es decir, estamos haciendo lo que antes llamábamos una previsión. Si ponemos valores decimales (1,0; 1,1; 1,2; 1,3 etc.), obtendremos muchos puntos y así podremos visibilizar mejor la tendencia. Es lo que aparece en el gráfico 23. Explicaré cómo hacer esto con una hoja de cálculo.

Por ejemplo, en la celda A1 escriba: 1,0

[1][,][0]

Después, en la celda A2 escriba: 1,1

[1][,][1]

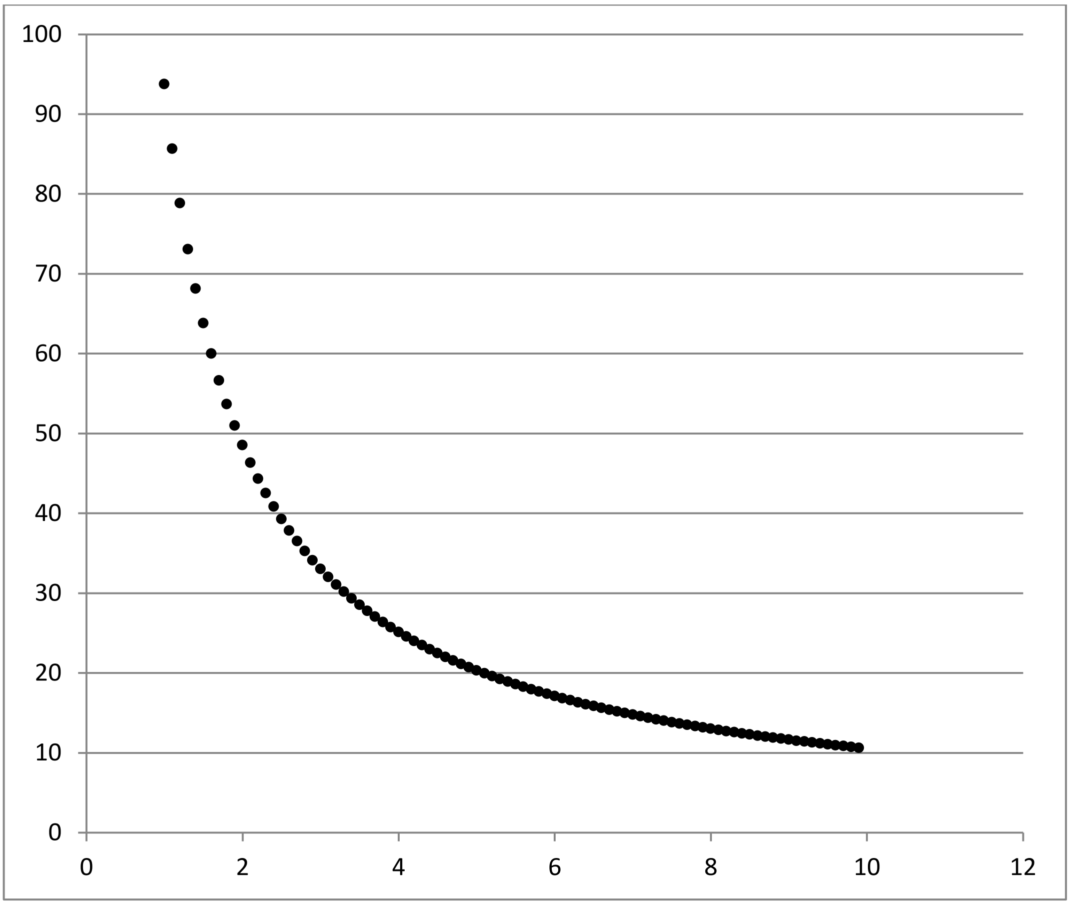
Después seleccione las dos celdas con el cursor. Aparecerá un recuadro con un punto marcado en el ángulo inferior derecho. Arrastre con el cursor ese punto hasta la celda A91 y observará como el programa va rellenando las celdas con valores que incrementan 0,1 cada vez. A continuación colocamos el cursor en la celda B1 y en la barra de fórmulas escribimos:

```
=93,766*POTENCIA(A1;-0,95)
```

Y vuelve a arrastrar el punto que hay en la parte inferior derecha de la selección de la celda B1 hasta la celda B91.

A continuación seleccione los datos de ambas columnas (A y B, en el ejemplo) y dígale al programa que inserte un gráfico de dispersión o de burbujas. El resultado es el gráfico 23.

[Gráfico 23]



Aquí ya tenemos los puntos (las moscas de Descartes o los bailarines del ejemplo) perfectamente ordenados. A continuación podemos, incluso, dar una forma matemática a nuestro teorema:

$$y = 93{,}766\, x^{-0{,}95}$$

Es lo mismo que:

$$y = \frac{93{,}766}{x^{0{,}95}}$$

O también es igual que:

$$y(x^{0{,}95}) = 93{,}766$$

O si queremos un enunciado más general:

Bajo rendimiento × Inversión pública = Constante

Que sería un teorema más simplificado.[34]

---

[34] Un caso de R bajo por valores discrepantes y la posibilidad de agrupar los datos, se encuentra aquí: https://www.fundacionexe.org/documents/demos009.pdf , pp. 56 y ss. (en catalán)

16.- UNA SALA DE BAILE CON VARIAS PISTAS

Hasta ahora hemos elaborado parejas de variables, imaginándonos que eran bailarines en una pista de baile. Pero ¿qué podemos hacer si tenemos más variables? Es decir, ¿cómo podemos aclararnos en una sala de baile donde suena la música de varias pistas al mismo tiempo? Ofreceremos dos posibilidades en este epígrafe y una más en el siguiente.

En primer lugar podemos proceder de la manera como hemos hecho hasta ahora, a saber, analizando solo parejas de variables, de manera que se simplifica el problema. Si queremos considerar tres variables, siempre podemos combinar dos de ellas (por ejemplo, haciendo su producto) y luego el resultado de la combinación relacionarlo con una tercera variable, y así sucesivamente. Si procede así, siempre hay que justificar por qué combinó dos variables, lo cual puede tener un cierto grado de arbitrariedad.

En segundo lugar, es posible hacer una matriz de correlaciones (lo que resulta más recomendable que el procedimiento anterior). Cuando tenemos más de dos variables, podemos hacer una matriz de correlaciones por pares de variables. Imaginemos que tenemos cinco variables V, W, X, Y y Z. Entonces podemos establecer la matriz de correlaciones de la tabla 6.

[Tabla 6]

	V	W	X	Y	Z
V	R_{VV}	R_{VW}	R_{VX}	R_{VY}	R_{VZ}
W	R_{WV}	R_{WW}	R_{WX}	R_{WY}	R_{WZ}
X	R_{XV}	R_{XW}	R_{XX}	R_{XY}	R_{XZ}
Y	R_{YV}	R_{YW}	R_{YX}	R_{YY}	R_{YZ}
Z	R_{ZV}	R_{ZW}	R_{ZX}	R_{ZY}	R_{ZZ}

En realidad, hay que hacer menos trabajo.

Sabemos que la correlación de una variable con ella misma siempre es 1, por lo que no es preciso calcular R_{VV}, R_{WW}, R_{XX}, etc., es decir, los valores de la diagonal.

Como tampoco importa el orden de las variables[35] (como ya se ha explicado, $R_{WV} = R_{VW}$, $R_{XV} = R_{VX}$, etc.) entonces solo es necesario calcular los valores a un lado de la diagonal. Por tanto, los cálculos que precisa la tabla 6 se pueden reducir a los que recoge la tabla 7:

---

[35] Recuérdese que al utilizar la hoja de cálculo teníamos que identificar una «matriz» con una variable y la otra «matriz» con la otra variable, pero el orden es indistinto. La palabra «matriz» aquí y allí no significan obviamente lo mismo.

[Tabla 7]

	V	W	X	Y	Z
V		R_{VW}	R_{VX}	R_{VY}	R_{VZ}
W			R_{WX}	R_{WY}	R_{WZ}
X				R_{XY}	R_{XZ}
Y					R_{YZ}
Z					

En realidad, en una matriz de correlaciones de N variables solo hemos de calcular M correlaciones, según la fórmula sencilla: $M = [N(N-1)]/2$. Las correlaciones que se precisa calcular en una matriz, en relación con el número de variables, se recogen en la tabla 8.

[Tabla 8]

Variables	Correlaciones
2	1
3	3
4	6
5	10
6	15
7	21
8	28
9	36
10	45

Variables	Correlaciones
11	55
12	66
13	78
14	91
15	105
16	120
17	136
18	153
19	171

Con un poco de experiencia con las fórmulas de las hojas de cálculo, esto resulta más fácil de lo que parece.

A continuación veremos un ejemplo.

Partimos de la tabla 8, donde se recogen datos sobre abandono educativo, puntuaciones de PISA, expectativa de escolarización, porcentaje de personas en educación superior o en aprendizaje permanente, inversión pública en educación y desigualdad social. Hemos depurado aquellos países que, como ya vimos, presentaban datos discordantes.

[Tabla 9]

	Abandono educativo temprano (%)	Puntuación Matemáticas (PISA 2012)	Puntuación Lectura (PISA 2012)	Puntuación Ciencias (PISA 2012)	Expectativa de escolarización (años)	Porcentaje Educación Superior (30-34 años) (%)	Aprendizaje permanente (%)	Inversión pública como porcentaje PIB (%)	Desigualdad social (ingresos quintil superior/quintil inferior)
Alemania	10,6	514	508	524	18,2	32,0	7,9	4,98	4,3
Austria	7,6	506	490	506	17,2	26,3	14,1	5,80	4,2
Bélgica	12,0	515	509	505	19,6	43,9	6,6	6,55	4,0
Chequia	5,5	499	493	508	18,1	25,6	10,8	4,51	3,5
Dinamarca	9,1	500	496	498	19,8	43,0	31,6	8,75	4,5
Eslovaquia	5,3	482	463	471	16,4	23,7	3,1	4,06	3,7
Eslovenia	4,4	501	481	514	18,5	39,2	13,8	5,68	3,4
España	24,9	484	488	496	17,9	40,1	10,7	4,82	6,5
Estonia	10,5	521	516	541	18,1	39,1	12,9	5,16	5,4
Finlandia	8,9	519	524	545	20,5	45,8	24,5	6,76	3,7
Francia	11,6	495	505	499	16,5	43,6	5,7	5,68	4,5
Holanda	8,8	523	511	522	19,1	42,2	16,5	5,93	3,6
Hungría	11,5	477	488	494	17,7	29,9	2,8	4,71	4,0
Irlanda	9,7	501	523	522	17,5	51,1	7,1	6,15	4,7
Italia	17,6	485	490	494	17,1	21,7	6,6	4,29	5,5
Letonia	10,6	491	489	502	17,9	37,2	6,9	4,93	6,5
Lituania	6,5	479	477	496	18,9	48,6	5,2	5,17	5,3
Luxemburgo	8,1	490	488	491	15,1	49,6	13,9	3,15	4,1
Polonia	5,7	518	518	526	18,3	39,1	4,5	4,94	4,9
Portugal	20,8	487	488	489	18,0	27,2	10,6	5,27	5,8
Reino Unido	13,6	494	499	514	16,6	47,1	15,8	5,88	5,0

Fuente: Eurostat. Datos referidos a 2012 o años anteriores en caso de ausencia del dato. Ahora Holanda se denomina Países Bajos.

A continuación, utilizando la hoja de cálculo (Excel), calculamos los coeficientes de correlación (R) entre pares de variables. El resultado se recoge en la tabla 10.

[Tabla 10]

	Abandono	Matemáticas	Lectura	Ciencias	Expectativa	Superior	Aprendizaje	Inversión	Desigualdad
Abandono		-0,31	-0,01	-0,21	-0,08	-0,12	-0,05	-0,05	0,68
Matemáticas			0,78	0,83	0,49	0,24	0,36	0,40	-0,33
Lectura				0,84	0,39	0,44	0,24	0,41	-0,07
Ciencias					0,48	0,36	0,33	0,34	-0,14
Expectativa						0,17	0,42	0,68	-0,12
Superior							0,30	0,36	0,02
Aprendizaje								0,66	-0,19
Inversión									-0,13
Desigualdad									

He sombreado las celdas que presentan valores superiores a 2/3 con un color gris oscuro y aquellas otras que están entre 1/3 y 2/3 con color gris claro. De esta manera «visual» podemos observar agrupaciones de variables.

En este ejemplo, por un lado, la desigualdad social correlaciona con el abandono educativo; por otro, la inversión se relaciona con la expectativa y el aprendizaje, y en un grado menor, con el rendimiento educativo. Es lógico que los resultados de las diversas materias de PISA correlacionen entre sí y, como se puede ver, en un grado más débil, con la expectativa de escolarización, educación superior, aprendizaje permanente e inversión.

Una vez que observamos las agrupaciones y la intensidad de la correlación, podemos proseguir con análisis más detallados. Dado que, como se ha explicado, el coeficiente de correlación es muy sensible a valores discrepantes, es bueno realizar las matrices de correlaciones una vez hayamos procedido a una cierta depuración de los valores (como he hecho aquí).

17.- UN PROGRAMA MÁS SOFISTICADO

Al principio de este texto hablé del criterio de usar solo programas habituales. Ahora haré una excepción, comentando una posibilidad de analizar varias variables con un programa más sofisticado.

Para analizar varias variables podemos hacer un análisis factorial, que también permite establecer relaciones numéricas entre diversas variables. No nos detendremos mucho en esta técnica porque requiere el programa especializado SPSS, que es el paquete estadístico de IBM (u otro similar). Haré una pequeña descripción y pondré de manera sucinta un ejemplo. El lector interesado en profundizar podrá encontrar prácticos tutoriales en Internet.

Lo que hace el análisis factorial es construir una serie de variables ficticias (llamadas «componentes»). El componente 1 sería la variable que mejor reproduciría la variabilidad del conjunto de las variables reales. El componente 2 sería otra variable ficticia que pretendería explicar la variabilidad no explicada por el componente 1 y así hasta que se explique el 100% de la variabilidad. Es decir, es como si el programa se inventara un bailarín («componente») que siguiera los movimientos de la mayoría de personas que danzan en una pista y pudiera medir qué porcentaje de los movimientos recoge. Después se inventara un nuevo bailarín para aquella parte no recogida, y así sucesivamente. El análisis factorial establece además el coeficiente de correlación entre los componentes y las variables reales, de manera que podemos saber qué variable se asemeja más en su movimiento al componente (al bailarín).

En la tablas 11 y 12 se recoge un ejemplo de análisis factorial con alguna de las variables indicadas anteriormente.

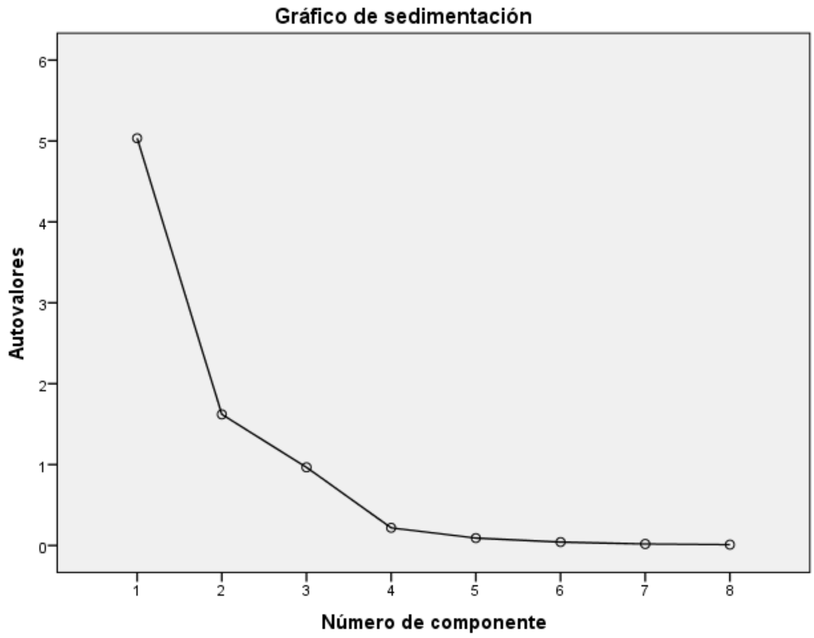
[Tabla 11]

Varianza total explicada		
	Sumas de las saturaciones al cuadrado de la extracción	
Componente	% de la varianza	% acumulado
1	62,930	62,930
2	20,255	83,185
3		

[Tabla 12]

Matriz de componentes		
	Componente	
	1	2
Puntuación PISA matemáticas 2012	0,917	
Bajo rendimiento PISA matemáticas 2012	-,0941	
Puntuación PISA lengua 2012	0,904	
Puntuación PISA ciencias 2012	0,909	
Expectativa de escolarización	0,651	
Porcentaje de aprendizaje permanente hombres	0,745	0,632
Porcentaje de aprendizaje permanente mujeres	0,700	0,625
Gasto anual por estudiante	0,442	0,663

[Gráfico 24]



En el ejemplo citado, el análisis factorial permite suponer un componente 1, que sería capaz de explicar el 62,9% de la variabilidad del grupo y que tendría una correlación de -0,941 con la variable «Bajo rendimiento PISA matemáticas 2012», es decir, esta variable sería la que mejor representaría a buena parte de los «bailarines».

18.- LAS FRECUENCIAS ESPERADAS

En muchos informes sobre educación se habla de las puntuaciones o de las frecuencias de un acontecimiento. Es lo que sucede cuando, por ejemplo, se dan las notas del estudiantado. Pero, ¿cómo saber si una puntuación es relativamente alta o baja? Este es el problema de la frecuencia esperada.

Comenzaremos con los datos de la tabla 13, donde aparecen los datos observados (aquí son datos; si fueran frecuencias hablaríamos de frecuencias observadas)

Para calcular las datos esperados procederemos a efectuar la suma de los valores de las filas y la suma de los valores de las columnas, así como la suma total, de las filas y de las columnas. En la tabla 13, estas sumas aparecen en la última columna y en la última fila.

[Tabla 13]

	Matemáticas	Lectura	Ciencias	Suma filas
Alemania	514	508	524	1546
Austria	506	490	506	1502
Bélgica	515	509	505	1529
Chequia	499	493	508	1500
Dinamarca	500	496	498	1494
Eslovaquia	482	463	471	1416
Eslovenia	501	481	514	1496
España	484	488	496	1468
Estonia	521	516	541	1578
Finlandia	519	524	545	1588
Francia	495	505	499	1499
Holanda	523	511	522	1556
Hungría	477	488	494	1459
Irlanda	501	523	522	1546
Italia	485	490	494	1469
Letonia	491	489	502	1482
Lituania	479	477	496	1452
Luxemburgo	490	488	491	1469
Polonia	518	518	526	1562
Portugal	487	488	489	1464
Reino Unido	494	499	514	1507
Suma columnas	10481	10444	10657	31582

Fuente: OCDE (PISA).

A continuación se calculan los datos esperados. Para cada celda hay que obtener el total de su fila multiplicado por el total de su columna, y el resultado dividirlo por el total global (la celda gris). Así, por ejemplo, el dato esperado de la primera celda (Alemania, matemáticas) es: $(1546 \times 10481)/31583$. El resultado es: 513,07 (redondeando 513). En la tabla 14 se ponen los datos esperadas. Naturalmente, la suma de las filas y de las columnas presentará los mismos valores que en la tabla anterior.

[Tabla 14]

	Matemáticas	Lectura	Ciencias	Suma filas
Alemania	513	511	522	1546
Austria	498	497	507	1502
Bélgica	507	506	516	1529
Chequia	498	496	506	1500
Dinamarca	496	494	504	1494
Eslovaquia	470	468	478	1416
Eslovenia	496	495	505	1496
España	487	485	495	1468
Estonia	524	522	532	1578
Finlandia	527	525	536	1588
Francia	497	496	506	1499
Holanda	516	515	525	1556
Hungría	484	482	492	1459
Irlanda	513	511	522	1546
Italia	488	486	496	1469
Letonia	492	490	500	1482
Lituania	482	480	490	1452
Luxemburgo	488	486	496	1469
Polonia	518	517	527	1562
Portugal	486	484	494	1464
Reino Unido	500	498	509	1507
Suma	10481	10444	10657	31582

Fuente: Elaboración de la tabla 13.

Por último, se resta el dato esperado del dato observado. Los resultados aparecen en la tabla 15. En esta tabla pueden observarse claramente aquellos datos que son más elevados que los datos esperados (valores positivos) y aquellos que son inferiores a los datos esperados (valores negativos). Así, por ejemplo, Eslovaquia obtiene una alta puntuación en matemáticas o Irlanda en Lectura, muy por encima de lo esperado; mientras que los resultados de matemáticas de Irlanda o de Lectura en Eslovenia están por debajo de lo esperado. A partir de estos datos, podemos indagar las causas de tales desviaciones.

(Aquí hemos eliminado la fila y la columna de las sumas, que lógicamente son ceros)

[Tabla 15]

	Matemáticas	Lectura	Ciencias
Alemania	1	-3	2
Austria	8	-7	-1
Bélgica	8	3	-11
Chequia	1	-3	2
Dinamarca	4	2	-6
Eslovaquia	12	-5	-7
Eslovenia	5	-14	9
España	-3	3	1
Estonia	-3	-6	9
Finlandia	-8	-1	9
Francia	-2	9	-7
Holanda	7	-4	-3
Hungría	-7	6	2
Irlanda	-12	12	0
Italia	-3	4	-2
Letónia	-1	-1	2
Lituania	-3	-3	6
Luxemburgo	2	2	-5
Polonia	0	1	-1
Portugal	1	4	-5
Reino Unido	-6	1	5

Una aplicación de este procedimiento es comparar las notas de una evaluación del estudiantado de un grupo. Otra es utilizarlo en los calculos de significatividad.

19.- El significado de la significatividad

Anteriormente afirmábamos que la «causalidad» es una «palabra mayor», y que podemos acercarnos a ella con «palabras menores». Una de ellas es «correlación», como hemos visto; otra es «significatividad» (estadística). Tenemos un procedimiento estadístico para comprobar si una hipótesis es «significativa» o no lo es, a saber, el cálculo de X^2 (aquí X no es una equis mayúscula, sino la letra griega «chi», que se pronuncia «ji» en castellano, o «qui» en portugués, por ello a este estadístico se le denomina habitualmente «ji cuadrado»). Este cálculo es relativamente fácil y se puede explicar con un ejemplo.

En la tabla 16 tenemos unos datos de rendimiento en Lengua correspondientes a estudiantes del sistema educativo español (pruebas PISA de 2015). Haremos la suposición de considerar que el porcentaje multiplicado por 10 (por lo que no cambian las proporciones) corresponde a individuos y que hay igual número de mujeres que de hombres (lo que sería coherente con una buena construcción de la muestra). Estas son las frecuencias observadas.

[Tabla 16]

	Bajo rendimiento	Entre bajo y alto rendimiento	Alto rendimiento	Total
Mujeres	128	809	63	1000
Hombres	196	757	47	1000
Total	324	1566	110	2000

Nos preguntamos: ¿influye el sexo en el rendimiento de lengua? El hecho de que no haya influencia de la variable (sexo), y que por lo tanto la variación que presenta la tabla se deba al azar, se denomina hipótesis nula (H_0). Lo contrario (en este caso, que sí que tiene influencia el sexo en el rendimiento de lengua) recibe el nombre de hipótesis alternativa.

En primer lugar, calcularemos la tabla de frecuencias esperadas. Como vimos en el epígrafe anterior, la frecuencia esperada en una celda es el producto del total de la fila por el total de la columna donde se encuentre esa celda, dividido por el total general. Así por ejemplo, en el caso de la celda de la fila Mujeres y la columna Bajo rendimiento, que presenta el valor 128, la frecuencia esperada sería (1000*324)/2000 = 162. De este modo confeccionaríamos la tabla de frecuencias esperadas, que sería la tabla 17:

[Tabla 17]

	Bajo rendimiento	Entre bajo y alto rendimiento	Alto rendimiento	
Mujeres	162	783	55	
Hombres	162	783	55	

En este caso particular, los valores de las filas de hombres y mujeres son los mismos porque sus totales son iguales, pero en otros casos puede no ser así. A continuación se calcula, para cada celda el valor correspondiente a esta fórmula:

$$\frac{(Frecuencia\ observada - Frecuencia\ esperada)^2}{Frecuencia\ esperada}$$

Y se suman todos esos valores, con lo que se obtiene el valor «calculado» de X^2. En nuestro caso:

$$X^2_{calculado} = 18{,}325$$

A continuación tenemos que establecer los grados de libertad (v), lo que corresponde a la fórmula:

$$v = (número\ de\ filas - 1)(número\ de\ columnas - 1) = (2-1)(3-2) = 1 \times\ 2 = 2$$

Posteriormente, establecemos el <u>margen de error</u>. En este caso adoptaremos un margen de error de $p = 0{,}05$ (o el 5%). A continuación consultaremos la tabla de X^2 (en el ANEXO I se ha reproducido una tabla) para establecer que X^2 corresponde en la tabla a ese valor. En nuestro caso, véase la fila segunda (v=2) y la columna 6ª (p=0,05), donde aparece el valor 5,9915. Por tanto:

$$X^2_{tabla} = 5{,}9915$$

La norma general es:

$$Si\ \ X^2_{calculado} < \ X^2_{tabla}\ entonces\ se\ acepta\ la\ H_0$$

$$Si\ \ X^2_{calculado}\ > \ X^2_{tabla}\ entonces\ se\ rechaza\ la\ H_0$$

En nuestro ejemplo, como:

$$(X^2_{calculado} = 18{,}325) > (X^2_{tabla} = 5{,}9915)\ entonces\ se\ rechaza\ la\ H_0$$

La conclusión es que el sexo sí que tiene significatividad (al 0,05) en el rendimiento de lengua.

Con todo, esta metodología de refutación de la hipótesis nula está siendo bastante criticada en ámbitos científicos[36] [37].

Si el lector o la lectura tiene más interés, puede leer el artículo: https://www.researchgate.net/profile/Valentin_Amrhein/publication/331908769_Scientists_rise_up_against_statistical_significance/links/5c9685a9299bf11169438d0a/Scientists-rise-up-against-statistical-significance.pdf

[36] Le debo a mi colega Vicent Garcia, de la Universitat de València, el haberme introducido en este campo de la crítica a la hipótesis nula.

[37] Mi colega Leopoldo Cabrera, de la Universidad de La Laguna, que amablemente revisó este texto, me aportó un ilustrativo ejemplo sobre el uso de X^2. No lo he incluido porque espero que él mismo lo publique en un amplio texto (en esta misma colección), que se aproveche de sus muchos años de magisterio en el mundo de la estadística y la investigación educativa, presentando esa disciplina de una manera fácil y práctica.

20.- UNA APLICACIÓN DE LA CORRELACIÓN Y LA PONDERACIÓN, PARA CALCULAR RIESGOS

Muchas investigaciones educativas son lo que se suele denominar indagaciones «end of pipe», es decir, investigaciones al final de la cañería. Por ejemplo, podemos obtener frecuencias del rendimiento del estudiantado o del abandono educativo temprano, pero estos son datos de cuando ya se ha producido el fenómeno. Es lo que en latín se decía «post festum», es decir, «después de la fiesta», aunque en muchas ocasiones los estudios sobre rendimiento o abandono no tengan nada de festivos. Por ello, usted puede pensar en investigaciones que, por así decir, se adelanten a los hechos, que, por ejemplo, permitan decir qué estudiantes están en riesgo de tener un mal (o un buen) rendimiento o efectuar un abandono educativo antes de que se produzca. Este conjunto de investigaciones podríamos denominarlo cálculo de riesgo. En el ámbito educativo podemos establecer diversas metodologías. En este epígrafe se presentará una y en el siguiente, otra.

Vamos a considerar un ejemplo de ponderación de promedios y cálculo de correlaciones, que, como se verá, es bastante sencillo.

Imaginemos que disponemos de los datos de n individuos (i_i), que representaremos como ($i_1, i_2, i_3 \dots\ i_n$) referidos a una serie de k variables (v_i), que representaremos como ($v_1, v_2, v_3 \dots\ v_k$), que podemos representar en una tabla (como la posterior tabla 19).

Una parte de esos individuos cumplen una determinada condición dicotómica (por ejemplo, han obtenido bajo rendimiento en una prueba). A estos m individuos los indicaremos como i'_i. Está claro que $n > m$ (es decir, todos los individuos son más que los que cumplen la condición; si no, no tendría sentido el estudio).

Pues bien, en primer lugar vamos a proceder a normalizar los datos de las variables, esto es, a expresarlos en el intérvalo 0-1.

Dado un dato cualquiera (x_j), su valor normalizado (x) se calcula con la fórmula siguiente, en la que también se tiene en cuenta el valor máximo de la serie (x_{max}) y el valor mínimo (x_{min}):

$$x = \frac{x_j - x_{min}}{x_{max} - x_{min}}$$

Esto es bastante sencillo de calcular con una hoja de cálculo. Basta poner los datos en una columna y escribir la fórmula en la siguiente, sustituyendo x_{max} y x_{min} por sus valores (o, fijando la celda en la que aparecen, pero esto exige un mayor conocimiento de la hoja de cálculo).

Entonces podemos representar los datos en la tabla siguiente (tabla 18).

[Tabla 18]

	$\boldsymbol{v_1}$	$\boldsymbol{v_2}$	**…**	$\boldsymbol{v_k}$	**Condición**
$\boldsymbol{i_1}$	$x_{1,1}$	$x_{1,2}$	…	$x_{1,k}$	Sí/No
$\boldsymbol{i_2}$	$x_{2,1}$	$x_{2,2}$	…	$x_{2,k}$	Sí/No
$\boldsymbol{i_3}$	$x_{3,1}$	$x_{3,2}$	…	$x_{3,k}$	Sí/No
…	…	…	…	…	…
$\boldsymbol{i'_1}$	$x'_{1,1}$	$x'_{1,2}$	…	$x'_{1,k}$	Sí/No
$\boldsymbol{i'_2}$	$x'_{2,1}$	$x'_{2,2}$	…	$x'_{2,k}$	Sí/No
$\boldsymbol{i'_3}$	$x'_{3,1}$	$x'_{3,2}$	…	$x'_{3,k}$	Sí/No
…	…	…	…	…	…
$\boldsymbol{i'_m}$	$x'_{m,1}$	$x'_{m,2}$	…	$x'_{m,k}$	Sí/No
…	…	…	…	…	…
$\boldsymbol{i_n}$	$x_{n,1}$	$x_{n,2}$	…	$x_{n,k}$	Sí/No

A continuación podemos calcular los promedios de cada variable para los individuos en general (i_i) y para los que cumplen la condición (i'_i), que denominaremos: $\overline{v_1}$, $\overline{v_2}$, $\overline{v_3}$ ... $\overline{v_k}$ y $\overline{v'_1}$, $\overline{v'_2}$, $\overline{v'_3}$ ... $\overline{v'_k}$.y podemos calcular las diferencias entre unas y otras (como valor absoluto, sin tener en cuenta el signo), de manera que:

$$d_i = |\overline{v}_i - \overline{v}'_i|$$

Si tenemos otro grupo de individuos y, que presenta los valores normalizados de las variables respectivas $(v_1, v_2, v_3 \ldots v_k)$ que ahora denominaremos $(w_1, w_2, w_3 \ldots w_k)$, podemos ponderarlas por las respectivas diferencias, lo que denominaremos (p_i)

$$p_i = w_i d_i$$

También estableceremos los valores del coeficiente de correlación (R) con la variable que consideremos más pertinente[38] (por ejemplo, v_1), lo que denominaremos R_i.

Pues bien, para un individuo y_i la suma de los respectivos productos $p_i R_i$ se presenta como un indicador del riesgo de cumplir la condición establecida. Ponderemos p_i por R_i para neutralizar (o reducir) la sobrerrepresentación de variables que correlacionen entre ellas.

[38] De nuevo aquí se nos cuela la «decisión», que tendremos que justificar en nuestro modelo.

Si dividimos esa suma por k tendría un valor entre 0 y 1, pero no es necesario si de lo que se trata es de analizar qué miembros de y están en situación de riesgo mayor de cumplir la condición. Si se hace la división, tendríamos una medida probabilística.

Naturalmente, al saber a posteriori qué miembros de y cumplen la condición, se ampliaría la base de individuos $(i + y)$, por lo que se tendría que recalcular v, w y R. Si tuviéramos definidas las variables de un nuevo grupo de individuos, el cálculo de riesgo estaría más perfeccionado, y así sucesivamente. De este modo procede, por ejemplo, la meteorología.

21.- OTRO PROCEDIMIENTO (MÁS SIMPLE) PARA CALCULAR RIESGOS

Supongamos un caso semejante al anterior, recogido en la tabla 18, pero más simple, en el que las variables estén dicotomizadas, esto es, los individuos solo puedan presentar dos valores, que podemos representar como 0 o 1. También en este caso distinguiremos una condición que presenta valores dicotómicos.

> Si disponemos de los datos en una hoja de cálculo, es muy fácil dicotomizarlos. Se puede hacer utilizando la función SI y asignando valores 0 o 1 según se cumpla o no la «Prueba lógica». Recuérdese que en esta prueba se pueden utilizar criterios como $\leq, \geq, <, >$, etc.

Se trata entonces de calcular la probabilidad de todas las combinaciones posibles. Si tenemos v variables, el número de combinaciones posibles corresponde a la fórmula:

$$2^v$$

Se trata de calcular la probabilidad de cada combinación, diviendo el número de individuos que cumplan la condición respecto de todos los individuos que pertenezcan a esa combinación.

La ventaja de este método es que nos permite trabajar con un gran número de individuos y realizar relativamente pocos cálculos (los cálculos a realizar dependen de las variables, no de los individuos). Además, se pueden utilizar las opciones de filtro de las hojas de cálculo para simplificar los cálculos.

Para elaborar las combinaciones posibles, podemos recurrir a una tabla del tipo siguiente. Por ejemplo, si se trata de 4 variables y si consideramos los valores 1 y 0, tendremos la tabla 19.

[Tabla 19]

v_1	v_2	v_3	v_4	Individuos que cumplen la condición en cada combinación ($casos\ favorables$)	Total de individuos de cada combinación ($casos\ posibles$)	Probabilidad: $\frac{casos\ favorables}{casos\ posibles}$
1	1	1	1			
1	1	1	0			
1	1	0	1			
1	1	0	0			
1	0	1	1			
1	0	1	0			
1	0	0	1			
1	0	0	0			
0	1	1	1			
0	1	1	0			
0	1	0	1			
0	1	0	0			
0	0	1	1			
0	0	1	0			
0	0	0	1			
0	0	0	0			

Se trata de contar en cada una de estas combinaciones la probabilidad asociada, es decir, el riesgo de presentar la condición, que es lo que hemos indicado en las columnas de la derecha.

No está de más recordar que la probabilidad de dos eventos no es la suma de la probabilidad de ambos. Recuérdese la fórmula general. Si tenemos dos eventos (A y B), la probabilidad (P) es:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

22.- HABLEMOS DE DESIGUALDAD

Ahora hablaremos de «desigualdad». De nuevo una de nuestras «palabras mayores». Por «desigualdad» no entendemos las meras diferencias, esto es, el hecho de que una persona es más alta que otra o pesa más, porque eso puede ser debido al azar o a circunstancias individuales. Hablamos de «desigualdad» cuando nos referimos a «estructuras», esto es, a configuraciones que van más allá de las circunstancias individuales y que están, por así decir, ancladas en la forma de nuestras relaciones sociales. Por ello, las desigualdades se repiten, son persistentes.

A continuación se comentará una de las maneras más inteligentes de medir la desigualdad social, el denominado índice de Gini, por dos razones: la primera es porque tiene muchas adaptaciones posibles a la investigación educativa; la segunda, porque es un modo muy elegante de pensar relaciones que se expresan no solo con curvas, sino también con superficies. Por ello, primero haremos una presentación habitual, la que el lector puede encontrar más o menos en cualquier manual o en cualquier enciclopedia, en papel o en línea. Ilustraremos esta presentación con un ejemplo propio. Después (en el epígrafe siguiente) pasaremos a una reflexión más profunda.

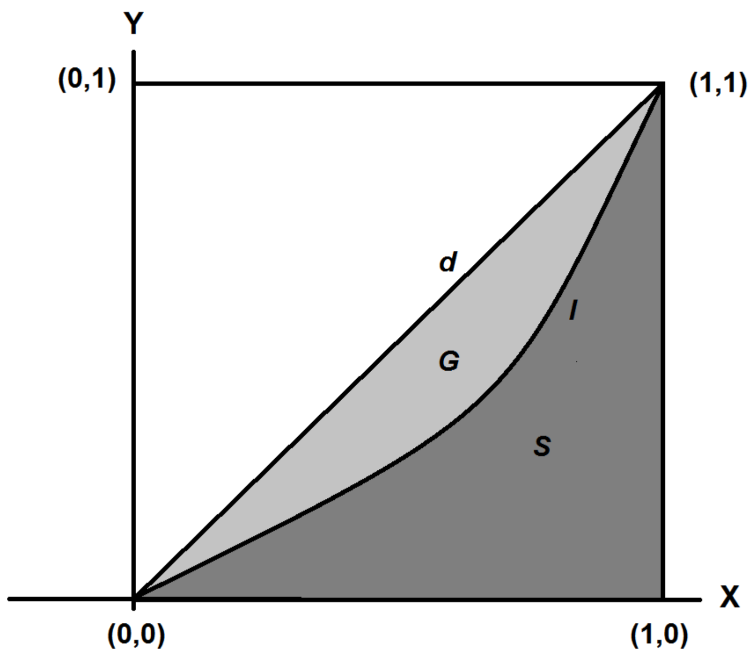
Como hemos dicho, los cuantiles establecen cortes en una muestra ordenada. Por ello, qué tan desigual sea un grupo se puede medir como una proporción de cuantiles. Para medir la desigualdad social generalmente se calcula la proporción entre el quintil superior y el quintil inferior de ingresos (lo que se abrevia S80/S20, por la palabra inglesa «share» y los puntos de corte, los cuantiles, del quintil superior –el que deja a su izquierda el 80% de la muestra– y el inferior –el que deja el 20%–). También se puede utilizar la comparación del decil superior y el inferior (lo que lógicamente se abreviaría S90/S10). Eurostat y otros organismos de estadística (por ejemplo, de la OCDE) publican las tablas de S80/S20 o S90/S10 de los distintos países y se pueden encontrar fácilmente en Internet.

Pero estas medidas tienen el defecto de no considerar qué pasa con los quintiles (o deciles) intermedios. Dos sociedades pueden tener los mismos ingresos en los quintiles superior e inferior pero diferencias notables en los quintiles intermedios. En ese caso, podríamos recurrir a una línea de tendencia que mostrara la pendiente de los quintiles (o los deciles), como hemos hecho anteriormente al hablar de cuartiles de rendimiento (cf. epígrafe 10.- Pendientes y ángulos). De ese modo podríamos también considerar el valor de los quintiles o los deciles intermedios. Pero pensamos esa pendiente como una línea recta (recuérdense los gráficos 3 y 4). ¿Qué sucedería si la pensáramos como una curva? Entonces estaríamos realizando una adaptación del llamado índice de Gini.

Imaginemos un país o un grupo social con una población que tiene unos ingresos (supongamos, anuales). Podemos ordenar esa muestra según el nivel de ingresos y representar la población acumulada (eje x) y los ingresos acumulados (eje y) en unos ejes cartesianos. Esto no plantea ninguna dificultad. Después trazaríamos la curva correspondiente. Para poder comparar esa curva con la de otro país podríamos representar no la población acumulada o los ingresos acumulados, sino su proporción respecto del total. A qué proporción de personas (en el eje x) le corresponde qué proporción de ingresos (en el eje y) (también hay otras razones matemáticas para trabajar con proporciones, que se explicarán más adelante).

En una situación de plena igualdad, si todos los individuos tuvieran exactamente los mismos ingresos, los porcentajes acumulados formarían una diagonal (o bisectriz). Por ejemplo, si la población fuera de 100 personas y cada una de ellas tuviera una remuneración de 10.000$ año, entonces la primera persona (el 1% de la muestra) aportaría una remuneración de 10.000$ (el 1% de los ingresos acumulados); dos personas, 20.000$; 3 personas, 30.000$, etc. etc. y por ello, la curva (recta) resultante sería la diagonal (d en el gráfico 31, la bisectriz que une los puntos (0,0) y (1,1)). Pero si en el grupo del ejemplo los ingresos son desiguales y los ordenamos de menor a mayor, la curva resultante estaría por debajo de la diagonal (l en el gráfico 25). A més desigualdad, más se alejaría la curva l de la diagonal d.

[Gráfico 25]



Pues bien, el índice de Gini (IG) se define como la proporción entre el área delimitada por la línea d y la curva l (marcada con la letra G, en gris claro en el gráfico) y toda el área por debajo de la línea d, es decir, el triángulo que une los puntos (0,0), (1,0) y (1,1), y que se compone del área bajo la curva (marcada con la letra S, en gris oscuro en el gráfico) y el área entre la línea d y la curva l. Es decir, el índice de Gini (IG) será:

$$IG = \frac{G}{G+S} = \frac{\frac{1}{2} - S}{\frac{1}{2}} = \frac{\frac{1}{2}}{\frac{1}{2}} - \frac{S}{\frac{1}{2}} = 1 - 2S$$

Por tanto, para calcular el índice de Gini (IG) necesitamos saber el valor de S, lo que no resulta fácil. Disponemos de dos procedimientos para calcular S. Los analizaremos a continuación con un cierto detalle, utilizando un ejemplo.

Nos planteamos, ¿Dónde es más desigual el gasto en educación superior, en España o en Portugal? Disponemos de los siguientes datos del gasto en educación superior por adulto y según quintiles (tabla 20).

[Tabla 20]

	Quintil 1º	Quintil 2º	Quintil 3º	Quintil 4º	Quintil 5º
España	11,214	43,476	34,384	59,457	159,462
Portugal	23,985	73,248	87,297	145,51	245,982

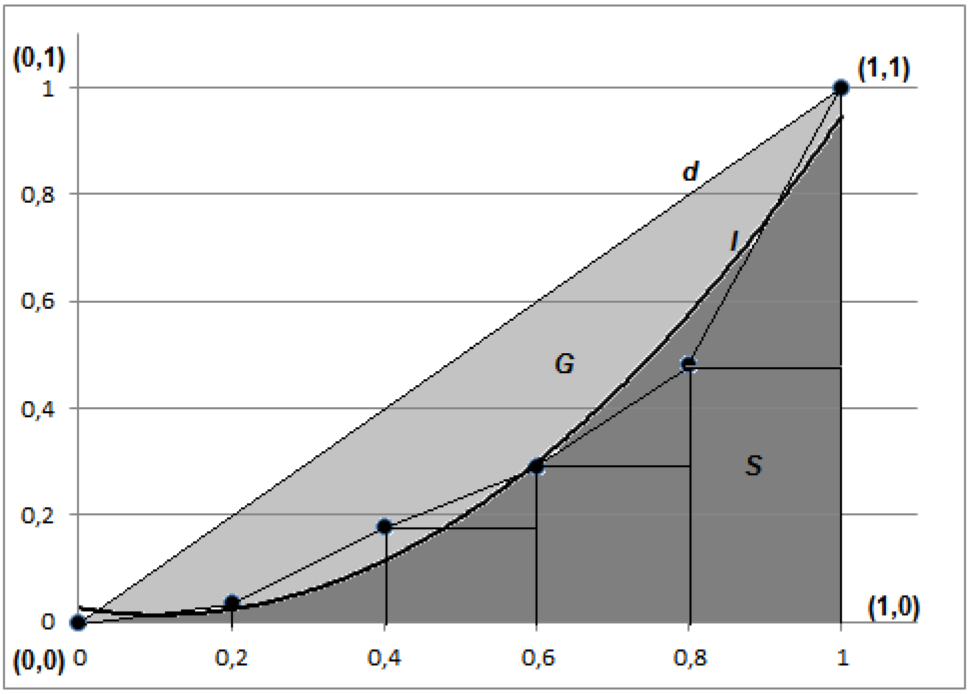
A continuación, calcularemos los valores acumulados y sus porcentajes (tabla 21)

[Tabla 21]

	Quintil 1º	Quintil 2º	Quintil 3º	Quintil 4º	Quintil 5º
España	0,0364	0,1776	0,2892	0,4823	1,0000
Portugal	0,0416	0,1688	0,3204	0,5730	1,0000

Representaremos los valores de España en el gráfico 26 para aclarar los procedimientos de cálculo.

[Gráfico 26]



Disponemos de dos procedimientos para calcular S (o mejor: para obtener aproximaciones satisfactorias), que podemos denominar de trapezoides (1º) y de área bajo curva (2º).

1º) Procedimiento de trapezoides.- Obsérvese el dibujo. El área S, en gris oscuro, es la suma de cinco trapezoides, o mejor, de un triángulo: el que está entre los puntos (0,0), (0, 0,2) y el valor del primer quintil, y cuatro trapezoides: el primero corresponde a los puntos (0,2, 0,4) y los valores del primer y el segundo quintil; el segundo, los puntos (0,4, 0,6) y los valores del segundo y tercer quintil, etc. Cada uno de estos trapezoides se compone de un rectángulo (de área base –es decir, 0,2– por altura –el valor del quintil anterior–) y un triángulo (semibase por la diferencia del quintil anterior y el posterior). Así pues, podemos sumar estas superficies y obtendremos el valor de S. Como el Índice de Gini $= 1 - 2S$, obtendremos los resultados de la tabla 22.

[Tabla 22]

	Área triángulo quintil 1º	Área trapezoide quintil 2º	Área trapezoide quintil 3º	Área trapezoide quintil 4º	Área trapezoide quintil 5º	Suma de áreas (S)	Índice de Gini (IG=1-2S)
España	0,003641	0,021398	0,046678	0,077146	0,148225	0,297088	0,405824
Portugal	0,004164	0,021044	0,048915	0,089332	0,157296	0,320751	0,358497

2) Procedimiento de área bajo curva.- El procedimiento del área bajo curva se basa en averiguar la función de la curva (l) y proceder a un cálculo de integrales. Como el cálculo de integrales no necesariamente forma parte de la enseñanza obligatoria, ofreceremos aquí una fórmula simplificada. Si el lector desea una demostración extensa de la fórmula, la encontrará en el **ANEXO II**.

Si disponemos de la función de la curva (l), como una función polinómica de orden 2 (es decir, como una ecuación de 2º grado), podemos calcular S del modo siguiente.

Si la ecuación es:

$$y = ax^2 + bx + c$$

El área de S es:

$$S = \frac{a}{3} + \frac{b}{2} + c$$

Y como $IG = 1 - 2S$, entonces tenemos la fórmula:

$$IG = 1 - \left(\frac{2}{3}a + b + 2c\right)$$

Así pues, tenemos los datos de la tabla 22 en una hoja de cálculo, añadimos también la pareja de datos (0,0) y solicitamos del programa que traze la línea de tendencia y la función polinómica de orden 2, que es la curva l que aparece en el gráfico 26, y después sustituimos los valores de a, b y c en la fórmula anterior.

El resultado se recoge en la tabla 23:

[Tabla 23]

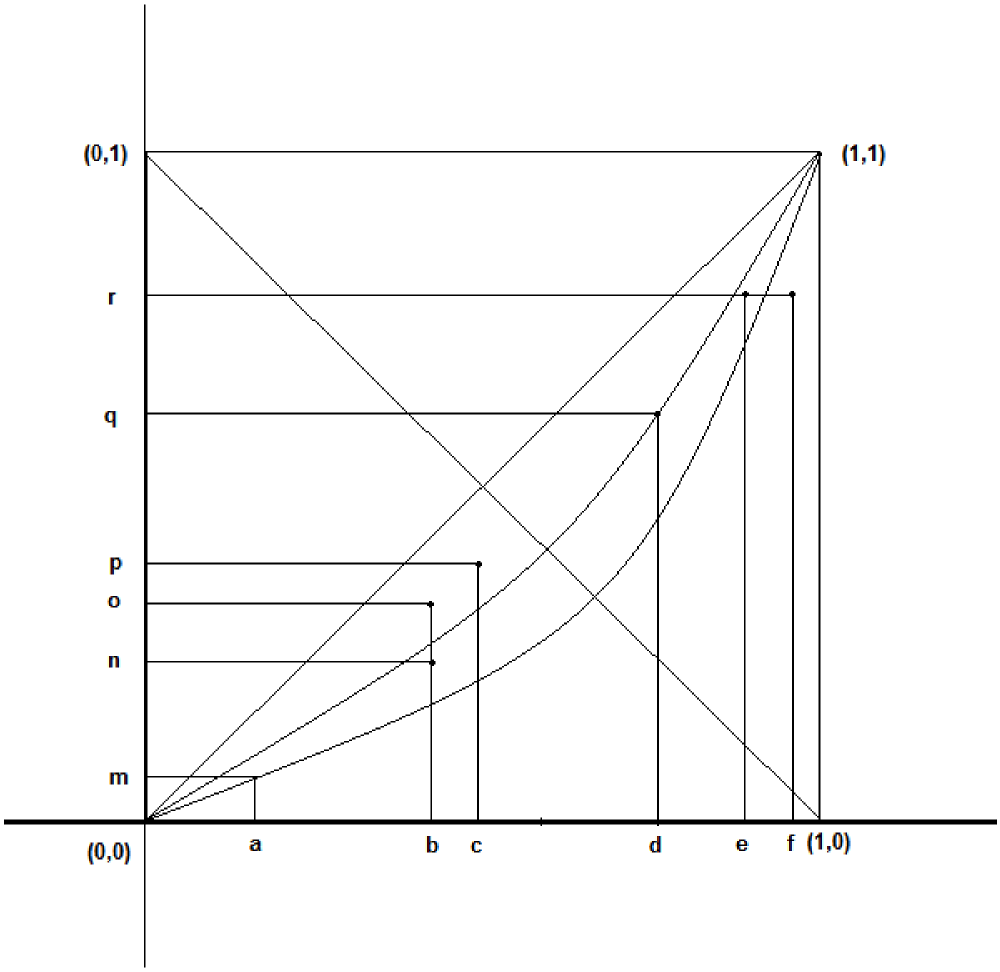
	Ecuación $ax^2 + bx + c$	$\frac{2}{3}a$	b	$2c$	Suma $\frac{2}{3}a + b + 2c$	Índice de Gini (IG=1-Suma)
España	$y = 1{,}1671x^2 - 0{,}2458x + 0{,}0259$	0,778067	-0,2458	0,0518	0,584067	0,415933
Portugal	$y = 1{,}0843x^2 - 0{,}1206x + 0{,}0134$	0,722667	-0,1206	0,0268	0,628867	0,371133

Como puede verse, los valores obtenidos por el primer procedimiento (España: 0,405; Portugal: 0,358) son muy parecidos a los obtenidos por el segundo procedimiento (España: 0,415; Portugal: 0,371). Por lo que podemos concluir, en este ejemplo, que hay más desigualdad en los gastos de enseñanza superior en España que en Portugal.

23.- Pensando la desigualdad como una curva

Observemos el gráfico 27 con atención. Es un gráfico semejante al gráfico 25, en el que se representa por una parte el porcentaje de población acumulada (eje x) y el porcentaje de riqueza acumulada (eje y). Los puntos corresponden a distintas sociedades <u>hipotéticas</u>. Vamos a ir comparándolos para preguntarnos cuál será la sociedad más desigual en cada caso. Recuerde que la <u>distancia</u> a la diagonal se entiende como <u>mayor desigualdad</u>.

[Gráfico 27]



Veamos en primer lugar los puntos (b,n) y (b,o), que corresponderían a dos sociedades distintas. ¿Cuál de las dos será más desigualitaria? La respuesta es fácil. Recuerde que la abscisa b corresponde a un porcentaje de población acumulada, luego el punto (b,n) se distancia más de la igualdad (de la diagonal) que el punto (b,o). Es decir, ese porcentaje de población, igual para los dos casos, tiene menor porcentaje de ingresos acumulados en el caso (b,n) que en el caso (b,o). Luego, está claro, el punto (b,n) corresponde a una sociedad más desigualitaria que el punto (b,o), y el punto (b,n) como se ve está más alejado de la diagonal, es decir, de la línea que une el punto (0,0) y el punto (1,1).

Ahora veamos los puntos (e,r) y (f,r) tienen el mismo valor en ordenadas (es decir, el mismo porcentaje de riqueza acumulada), pero distinto valor en abscisas. Naturalmente, la sociedad que corresponde al punto (e,r) será más igualitaria que la del punto (f,r) porque precisa menos población para llegar a la misma proporción de ingresos acumulados. Y también en este caso el punto (f,r) como se ve está más alejado de la diagonal, es decir, de la línea que une el punto (0,0) y el punto (1,1).

En estos cuatro casos, tomados dos a dos, no parece haber dudas de qué puntos corresponden a sociedades más igualitarias o más desigualitarias, porque coincide una abscisa o una ordenada (es decir, un valor de x o un valor de y) y el punto de la sociedad más desigualitaria está más alejado de la diagonal.

En tercer lugar, veámos el caso de los puntos (a,m), (c,p) y (d,q). ¿Cuál corresponde a una sociedad más igualitaria o a una sociedad más desigualitaria? Tal vez usted conteste un poco irreflexivamente que será más desigualitaria la sociedad que corresponda a un punto más alejado de la diagonal, porque eso es lo que hemos concluido en los casos anteriores. Pero esta respuesta <u>no es aceptable</u>, porque, recuérdese, que la representación de la desigualdad (gráfico 25) pasaba por los puntos (0,0) y (1,1), porque es obvio que con una proporción de población acumulada $x = 0$, entonces la proporción de riqueza acumulada tiene que ser $y = 0$, y al contrario, una proporción de riqueza acumulada $y = 1$, ha de corresponder a toda la población, y por tanto $x = 1$. Por eso, no es correcta la respuesta de «será más desigualitaria la sociedad que corresponda a un punto más alejado de la diagonal», porque eso es seguir pensando en términos de línias rectas paralelas (a la diagonal, es decir, distancias determinadas), y no en términos de líneas curvas que pasan por uno esos puntos y también por (0,0) y (1,1). Es por ello que hemos marcado dos curvas parabólicas que pasan por los puntos (a,m) y (d,q), para poder visibilizar mejor esa <u>curvatura de la desigualdad</u>.

24.- «LÍNEAS RECTAS Y LÍNEAS CURVAS, LO IMPORTANTE ES...»

Un viejo proverbio chino, muy citado en el debate político español, decía: «gato blanco, gato negro; lo importante es que cace ratones». Aquí también podríamos decir que lo importante es aprehender la desigualdad social, y eso lo podemos hacer, como hemos visto, con líneas rectas o curvas. Ahora que ya se ha explicado el Índice de Gini, podemos volver sobre un ejemplo anterior, el del bajo rendimiento según las pruebas PISA que se explicaba en el epígrafe 10, con el recurso a líneas de tendencia rectas y el cálculo de sus respectivas tangentes o arco tangentes. Ahora podemos hacer estos cálculos utilizando el Índice de Gini.

Sigamos el cálculo con detalle. Tomemos el primer país de la lista de la OCDE: Australia. La base de datos de PISA proporciona los porcentajes de bajo rendimiento en ciencias. Los datos se recogen en la tabla 24, donde se ha añadido también la suma de los porcentajes.

[Tabla 24]

	Cuartil 1º	Cuartil 2º	Cuartil 3º	Cuartil 4º	Suma
Australia	70,80%	80,80%	87,60%	93,30%	332,50%

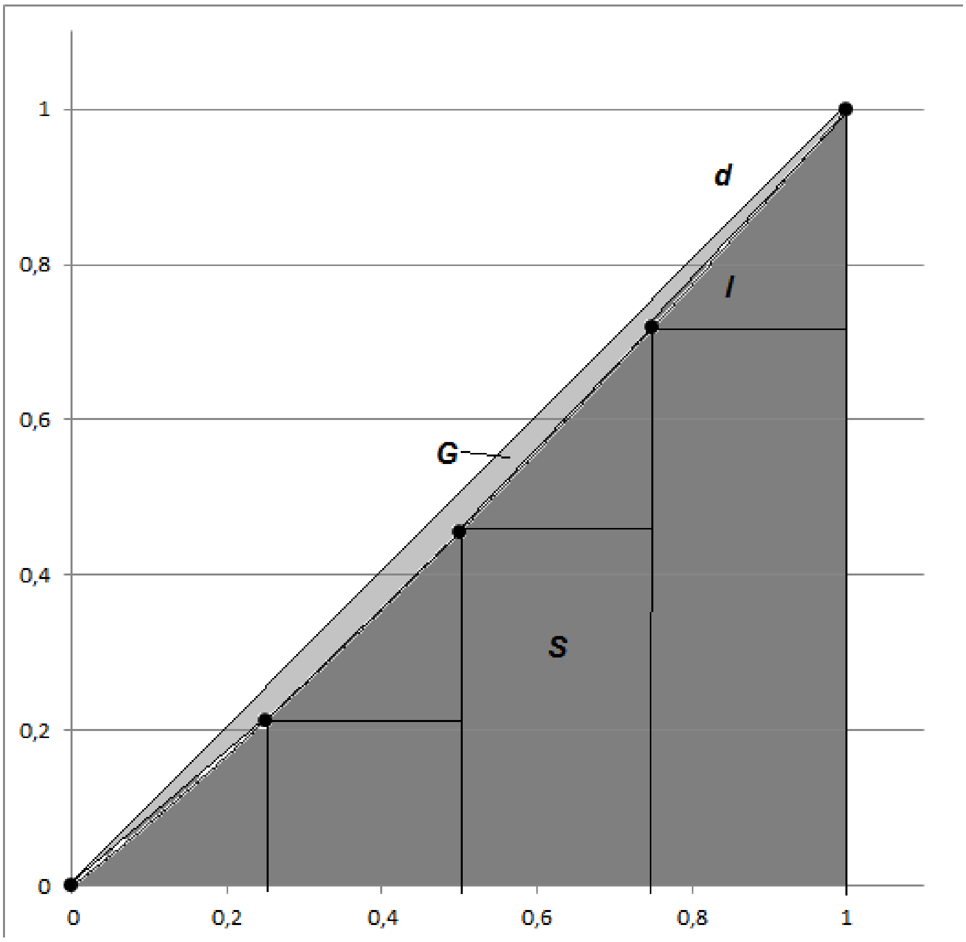
En realidad, daría lo mismo que consideráramos porcentajes o valores absolutos; se trataría de valores de y. A continuación se calcula el porcentaje de cada valor (70,80 etc., respecto de 332,50) y los sucesivos porcentajes acumulados. Se obtiene así la tabla 25.

[Tabla 25]

	Acumulado Cuartil 1º	Acumulado Cuartil 2º	Acumulado Cuartil 3º	Acumulado Cuartil 4º
Australia	0,21293	0,45594	0,719398	1

Podemos representar estos valores en un gráfico, para lo que hay que tener en cuenta que cada cuartil representa un 20% de la muestra (o 0,25 en términos decimales) y que hemos de añadir el par (0,0). El resultado es el gráfico 28.

[Gráfico 28]



Para calcular el área S procederemos de la manera siguiente. El área del primer triángulo es la base (0,25) por la altura (acumulado del cuartil 1º), dividida por 2. En el cuartil segundo tenemos un rectángulo, cuya área es base (0,25) por altura (acumulado del cuartil 1º), más un triángulo, cuya área la base (0,25) por la altura (acumulado del cuartil 2º menos acumulado del cuartil 1º) dividida por 2, y lo mismo para el cuartil tercero y el cuarto. El resultado del cálculo de estas áreas se presenta en la tabla 19, así como la suma de los cuatro cuartiles (S). Finalmente, como sabemos que el Índice de Gini se puede calcular a partir del valor de S com la fórmula: $= 1 - 2S$, podemos averiguar el valor del Índice, tal como se recoge en la tabla 26.

[Tabla 26]

	Área Cuartil 1º	Área Cuartil 2º	Área Cuartil 3º	Área Cuartil 4º	Suma áreas (S)	IG (=1-2S)
Australia	0,026617	0,083609	0,146917	0,214925	0,472068	0,055865

Por tanto, el Índice de Gini para Australia es: 0,055865 o, expresado como Coeficiente de Gini: 5,5865. Calculando país por país (lo que resulta bastante fácil si definimos una hoja de cálculo con todas las operaciones indicadas), obtenemos los resultados de las tablas 27 y 28.

[Tabla 27]

	No bajo rendimiento, según índice socioeconómico			
	Cuartil 1º	Cuartil 2º	Cuartil 3º	Cuartil 4º
Australia	70,80%	80,80%	87,60%	93,30%
Austria	64,90%	76,40%	85,20%	91,60%
Belgium	64,80%	77,10%	87,20%	94,10%
Canada	81,40%	88,30%	92,40%	94,60%
Chile	43,80%	65,00%	69,30%	84,30%
Czech Republic	63,50%	77,40%	84,20%	94,00%
Denmark	74,70%	81,30%	88,90%	93,20%
Estonia	86,50%	90,30%	92,40%	96,30%
Finland	80,30%	86,80%	92,00%	95,50%
France	60,10%	74,80%	86,40%	94,80%
Germany	72,10%	84,20%	88,20%	95,20%
Greece	50,20%	62,30%	72,20%	85,30%
Hungary	53,00%	72,40%	78,80%	92,00%
Iceland	66,70%	73,60%	77,50%	82,30%
Ireland	73,60%	82,70%	89,00%	93,70%
Israel	51,80%	65,90%	77,20%	82,20%
Italy	63,10%	76,80%	80,10%	88,50%
Japan	82,80%	90,40%	92,90%	96,60%
Korea	76,80%	83,50%	89,70%	93,00%
Latvia	75,00%	79,20%	86,00%	91,60%
Luxembourg	54,90%	70,70%	80,10%	92,50%
Mexico	34,80%	47,70%	57,40%	69,70%
Netherlands	69,80%	78,90%	85,10%	93,10%
Norway	71,50%	80,50%	85,20%	89,70%
Portugal	70,10%	81,20%	84,20%	95,50%
Slovak Republic	50,10%	68,20%	74,70%	86,50%
Slovenia	74,90%	82,50%	90,00%	93,70%
Spain	68,40%	79,20%	86,00%	94,00%
Sweden	66,40%	75,90%	85,60%	89,30%
Switzerland	67,30%	80,80%	85,30%	93,70%
Turkey	42,20%	51,20%	57,60%	72,10%

United Kingdom	74,30%	78,40%	87,60%	92,80%
United States	68,00%	76,20%	84,90%	91,40%
Algeria	25,20%	25,70%	28,50%	38,30%
Brazil	27,70%	38,60%	44,90%	65,40%
B-S-J-G (China)	68,90%	84,80%	87,30%	94,00%
Bulgaria	40,90%	56,10%	71,10%	83,00%
CABA (Argentina)	50,70%	76,90%	87,90%	93,70%
Colombia	34,90%	42,70%	53,90%	72,50%
Costa Rica	35,90%	46,30%	56,40%	76,20%
Croatia	64,00%	71,50%	77,50%	88,90%
Cyprus	43,10%	53,50%	60,90%	75,10%
Dominican Republic	3,30%	8,20%	13,90%	31,90%
FYROM	25,20%	34,40%	39,00%	51,20%
Georgia	32,10%	41,00%	55,90%	68,50%
Hong Kong (China)	85,90%	89,80%	92,30%	95,30%
Indonesia	28,90%	38,10%	45,60%	63,80%
Jordan	32,80%	47,50%	55,70%	67,10%
Kosovo	23,40%	28,30%	31,40%	47,30%
Lebanon	21,90%	32,30%	38,40%	57,80%
Lithuania	61,30%	71,20%	81,00%	88,40%
Macao (China)	89,90%	92,70%	92,00%	93,40%
Malta	50,40%	63,40%	72,40%	85,10%
Moldova	40,60%	55,90%	59,40%	75,80%
Montenegro	38,10%	44,90%	50,70%	63,70%
Peru	15,10%	35,50%	48,10%	67,60%
Qatar	33,70%	52,10%	59,90%	57,00%
Romania	43,90%	55,60%	65,80%	80,80%
Russia	72,90%	80,70%	86,90%	88,60%
Singapore	78,90%	90,80%	94,10%	97,90%
Chinese Taipei	76,90%	86,60%	90,90%	96,00%
Thailand	43,80%	46,70%	52,70%	71,60%
Trinidad and Tobago	41,40%	49,40%	58,00%	73,10%
Tunisia	21,90%	28,50%	34,50%	53,90%
United Arab Emirates	44,10%	56,70%	66,60%	66,90%
Uruguay	40,70%	53,00%	63,00%	81,00%
Viet Nam	90,60%	93,70%	94,70%	97,40%

[Tabla 28]

	Área del trapezoide correspondiente a cada cuartil				Suma	IG
	Cuartil 1º	Cuartil 2º	Cuartil 3º	Cuartil 4º	(S)	(1-2*S*)
Australia	0,026617	0,083609	0,146917	0,214925	0,472068	0,055865
Austria	0,025503	0,081028	0,144530	0,214005	0,465066	0,069868
Belgium	0,025062	0,079943	0,143487	0,213606	0,462098	0,075804
Canada	0,028525	0,087994	0,151318	0,216849	0,484686	0,030628
Chile	0,020865	0,072694	0,136671	0,209842	0,440072	0,119855
Czech Republic	0,024875	0,080069	0,143372	0,213178	0,461493	0,077013
Denmark	0,027618	0,085293	0,148218	0,215543	0,476671	0,046658
Estonia	0,029583	0,090048	0,152531	0,217066	0,489227	0,021546
Finland	0,028307	0,087211	0,150240	0,216335	0,482092	0,035815
France	0,023766	0,077112	0,140857	0,212512	0,454247	0,091506
Germany	0,026531	0,084045	0,147483	0,214969	0,473028	0,053945
Greece	0,023241	0,075324	0,137593	0,210509	0,446667	0,106667
Hungary	0,022367	0,075287	0,139095	0,211175	0,447924	0,104153
Iceland	0,027782	0,086221	0,149159	0,215720	0,478882	0,042236
Ireland	0,027139	0,084771	0,148083	0,215450	0,475442	0,049115
Israel	0,023367	0,076462	0,141014	0,212920	0,453762	0,092476
Italy	0,025567	0,082253	0,145827	0,214141	0,467788	0,064425
Japan	0,028536	0,088227	0,151399	0,216708	0,48487	0,030259
Korea	0,027988	0,086407	0,149526	0,216108	0,480029	0,039942
Latvia	0,028255	0,086347	0,148583	0,215491	0,478677	0,042646
Luxembourg	0,023013	0,075662	0,138875	0,211226	0,448776	0,102448
Mexico	0,020754	0,069955	0,132634	0,208433	0,431775	0,136450
Netherlands	0,026690	0,083550	0,146260	0,21440	0,470901	0,058198
Norway	0,027340	0,085462	0,148822	0,215701	0,477325	0,045350
Portugal	0,026473	0,08361	0,146073	0,213935	0,470091	0,059819
Slovak Republic	0,022406	0,075313	0,139222	0,211315	0,448256	0,103488
Slovenia	0,027448	0,085129	0,148344	0,215663	0,476583	0,046834
Spain	0,026099	0,082418	0,145452	0,214133	0,468101	0,063797
Sweden	0,026166	0,082243	0,145886	0,214809	0,469105	0,061791
Switzerland	0,025718	0,082314	0,145789	0,214193	0,468014	0,063971
Turkey	0,023644	0,075975	0,136934	0,209603	0,446156	0,107687
United Kingdom	0,027882	0,085185	0,147478	0,215176	0,475721	0,048559
United States	0,026521	0,082761	0,145593	0,214353	0,469228	0,061544
Algeria	0,026763	0,080820	0,138381	0,209325	0,455289	0,089422
Brazil	0,019606	0,066535	0,125637	0,203709	0,415487	0,169026
B-S-J-G (China)	0,025709	0,083060	0,147276	0,214925	0,470907	0,058060
Bulgaria	0,020360	0,068648	0,131969	0,208682	0,429659	0,140681

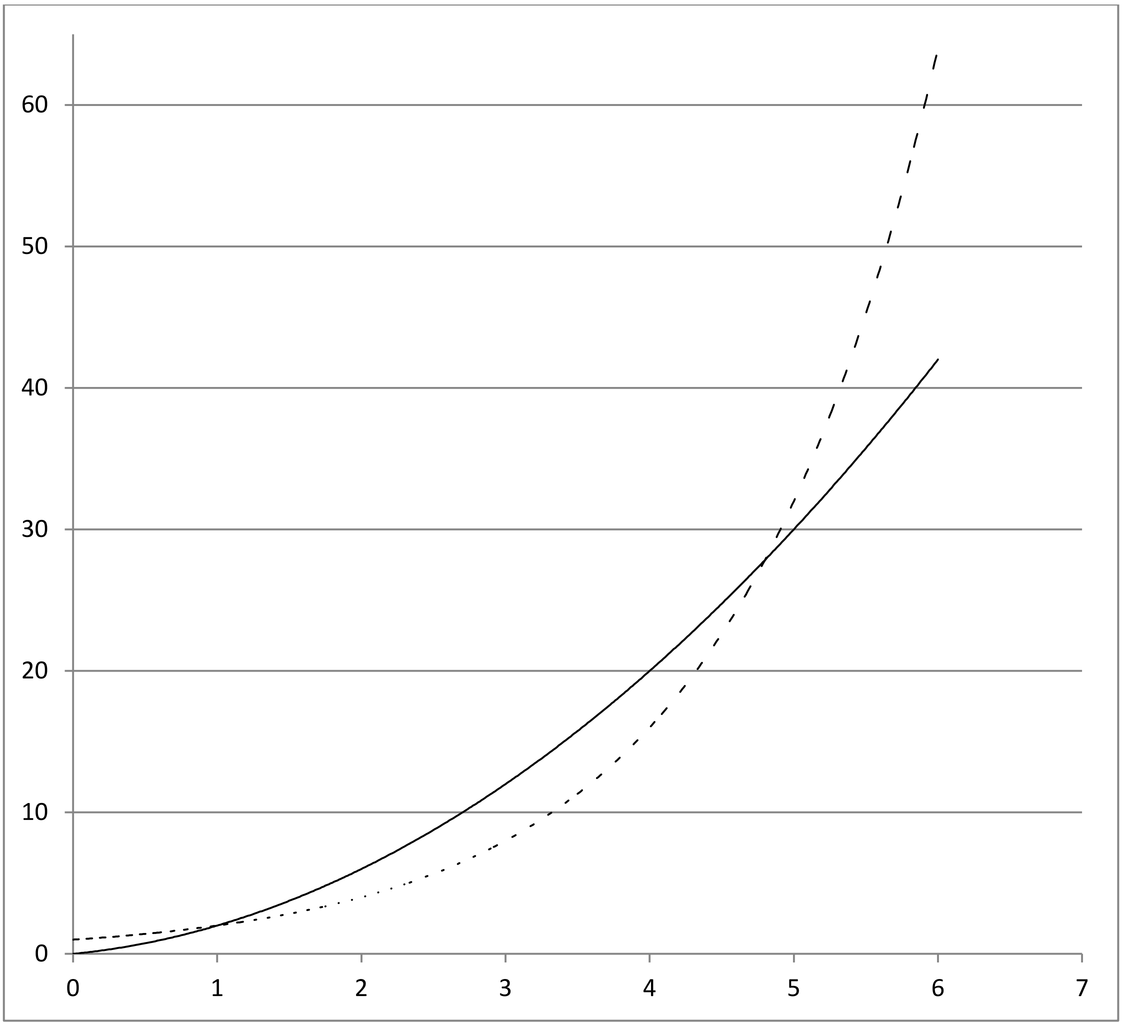
CABA (Argentina)	0,020496	0,072081	0,138705	0,212120	0,443402	0,113195
Colombia	0,021385	0,068934	0,128125	0,205576	0,424020	0,151961
Costa Rica	0,020892	0,068727	0,128492	0,205656	0,423766	0,152467
Croatia	0,026499	0,082602	0,144294	0,213191	0,466587	0,066827
Cyprus	0,023162	0,075075	0,136554	0,209641	0,444433	0,111135
Dominican Republic	0,007199	0,032286	0,080497	0,180410	0,300393	0,399215
FYROM	0,021028	0,070761	0,132009	0,207276	0,431075	0,137850
Georgia	0,020316	0,066582	0,127911	0,206646	0,421456	0,157089
Hong Kong (China)	0,029555	0,090008	0,152663	0,217210	0,489437	0,021126
Indonesia	0,020479	0,067956	0,127268	0,204790	0,420493	0,159014
Jordan	0,020187	0,069609	0,133124	0,208703	0,431622	0,136755
Kosovo	0,022431	0,071990	0,129218	0,204659	0,428298	0,143405
Lebanon	0,018201	0,063248	0,122008	0,201961	0,405419	0,189162
Lithuania	0,025381	0,080242	0,143259	0,213398	0,462281	0,075439
Macao (China)	0,030537	0,092561	0,155299	0,218274	0,496671	0,006658
Malta	0,023222	0,075654	0,138223	0,210791	0,447890	0,104220
Moldova	0,021903	0,073964	0,136167	0,209107	0,441142	0,117717
Montenegro	0,024126	0,076684	0,137221	0,209663	0,447695	0,104610
Peru	0,011350	0,049384	0,112222	0,199188	0,372144	0,255713
Qatar	0,020782	0,073693	0,142760	0,214850	0,452084	0,095831
Romania	0,022298	0,072836	0,134498	0,208960	0,438592	0,122816
Russia	0,027689	0,08603	0,149689	0,216348	0,479755	0,040489
Singapore	0,027267	0,085914	0,149813	0,216167	0,479161	0,041678
Chinese Taipei	0,027433	0,085759	0,149080	0,215753	0,478025	0,043950
Thailand	0,025489	0,078154	0,135999	0,208333	0,447975	0,104050
Trinidad and Tobago	0,023321	0,07447	0,134971	0,208822	0,441584	0,116832
Tunisia	0,019723	0,065112	0,121848	0,201459	0,408141	0,183718
United Arab Emirates	0,023528	0,077305	0,143086	0,214309	0,458227	0,083547
Uruguay	0,021403	0,070677	0,131679	0,207404	0,431163	0,137674
Viet Nam	0,030088	0,091293	0,153859	0,217654	0,492893	0,014214

Los siguientes epígrafes son un tanto sofisticados. Tratan de un caso particular, que es el establecimiento del Índice de Gini con un par de valores x_i,y_i. Por eso, si no es el caso de su investigación, puede prescindir de su lectura y pasar al epígrafe de conclusión.

25.- La desigualdad, con un par de valores y una curva parabólica...

En el gráfico 29 se muestran dos tipos de curvas: con trazo continuo, una curva parabólica, que corresponde a la función: $y = x^2 + x$; con trazo discontinuo, una curva exponencial, que corresponde a la función: $y = 2^x$. Es bueno recordar esta distinción para poder seguir las explicaciones que siguen.[39]

[Gráfico 29]



[39] La profesora de matemáticas Josefa Marín, docente en la Universitat Politècnica de València, atendió amablemente mis consultas y me ayudó a resolver algunos problemas de los fragmentos siguientes.

A continuación plantearemos un problema y explicaremos las maniobras geométricas que tenemos que hacer para fundamentar una respuesta.

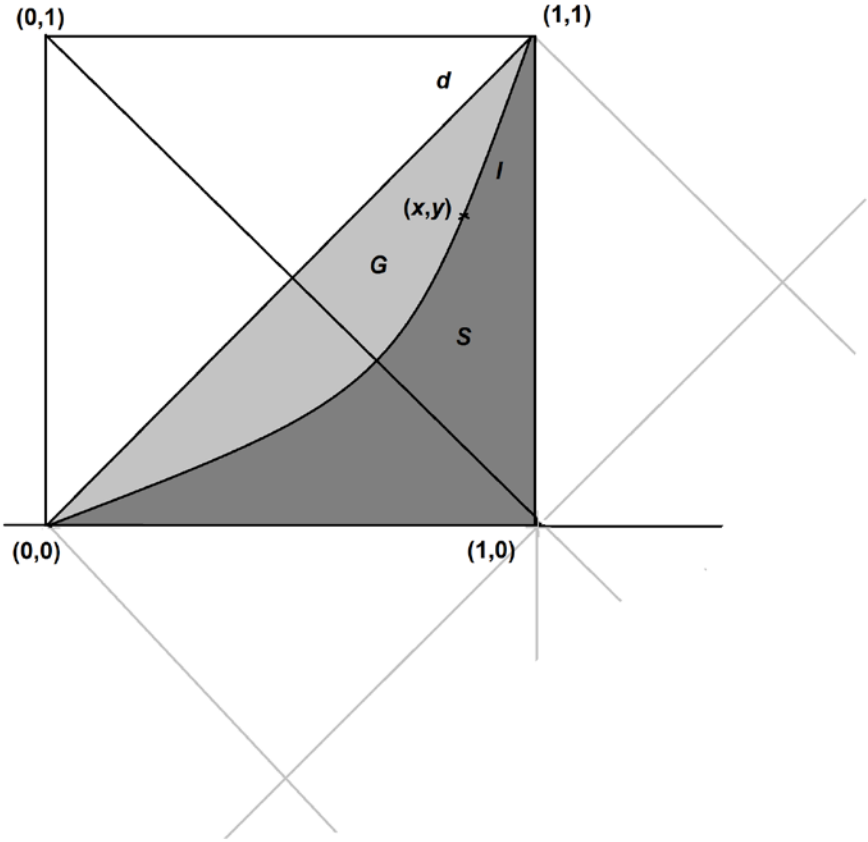
Observemos los datos reales de dos titulaciones en una universidad real: Derecho y Enfermería. En el curso 2013/14, un 20,3% de las personas matriculadas en la Facultad de Derecho eran nuevos estudiantes. Ese mismo curso, en la Facultad de Enfermería los nuevos estudiantes representaban el 30,1%. Ambos grados tienen una duración teórica de cuatro cursos. Así pues, cuatro cursos después, habían concluido el 20,00% de los nuevos estudiantes de Derecho y el 70,92% de los de Enfermería. ¿En cuál de las dos titulaciones el estudiantado «fluye» mejor?

A primera vista, uno se inclinaría a contestar que en Enfermería, pero, ¿cómo calcularlo de una manera precisa y, mejor todavía, poder asignar un índice que represente la falta de fluidez? Para contestar a esta cuestión realizaremos una adaptación del Índice de Gini, que se comenta a continuación.

En el gráfico 25 hemos introducido subrepticiamente un modelo para analizar la desigualdad: una curva que pasa por el punto en cuestión (x_i, y_i) y por los puntos (0,0) y (1,1), y que nos permite calcular una medida de desigualdad a la manera del Índice de Gini. De nuevo podemos recurrir a trazar una línea de tendencia (curva) parabólica que pase por los tres puntos y solicitarle al programa de cálculo que nos proporcione la ecuación de esa curva. Una manera de hacer esto es realizar una pequeña transformación geométrica: primero un desplazamiento en los ejes cartesianos y luego un giro. Intentaremos representar este doble movimiento con los gráficos siguientes, que exigen del lector un cierto esfuerzo de abstracción.

Comenzaremos por un gráfico igual que el anterior gráfico 5, en el que hemos introducido unas líneas complementarias para que se puedan visibilizar mejor las transformaciones que vamos a realizar. Recuérdese que lo que nos interesa de este gráfico son las superficies marcadas como G y S, porque la medida de la desigualdad, el Índice de Gini, la hemos definido como $IG = \frac{G}{G+S} = 1 - 2S$. Además el lector tiene que fijar su atención en los cuatro puntos fundamentales del gráfico: los puntos (0,0), (1,0), (0,1), (1,1) y (x,y).

[Gráfico 30]

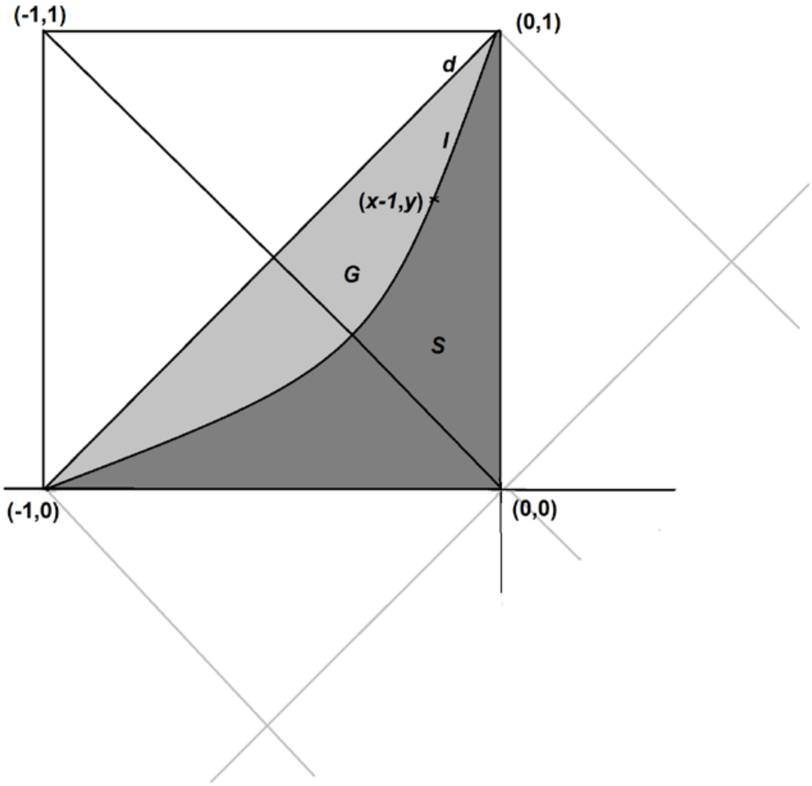


A continuación vamos a proceder a una traslación de los ejes de coordenadas, que serán desplazados una unidad hacia la derecha. De este modo, todos los puntos del gráfico tienen unas nuevas coordenadas. La transformación se explica en la tabla 29 y se representa en el gráfico 31.

[Tabla 29]

Coordenadas gráfico 35	Coordenadas gráfico 30, después desplazamiento
(0,0)	(-1,0)
(0,1)	(-1,1)
(1,0)	(0,0)
(1,1)	(0,1)
(*x*, *y*)	(*x*-1, *y*)

[Gráfico 31]

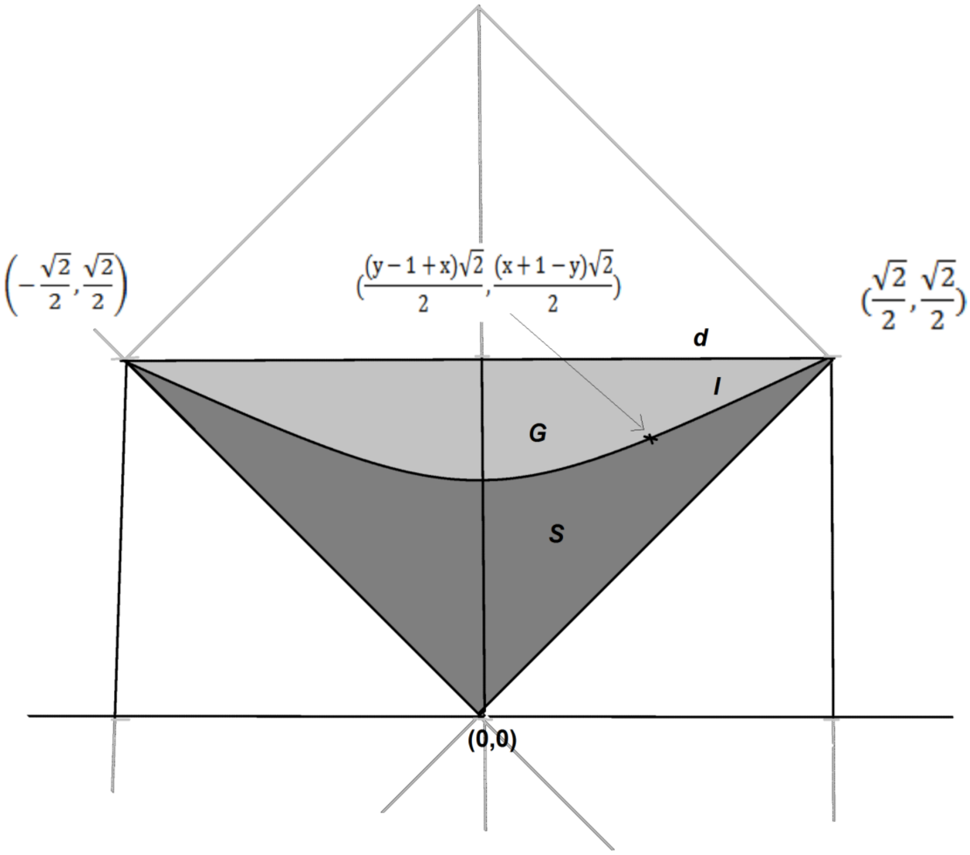


A continuación, se efectúa un giro de 45º (en el sentido de las agujas del reloj) de los ejes cartesianos. Naturalmente, los cuatro puntos tendrán entonces nuevas coordenadas, aunque no cambien las áreas G y S. Las nuevas coordenadas se recogen en la tabla 30 y la representación después del giro en el gráfico 32 (que a su vez ha sido girado 45º en el sentido contrario a las agujas del reloj, para favorecer la contemplación por parte del lector).

[Tabla 30]

Coordenadas gráfico 35	Coordenadas gráfico 30, después desplazamiento	Coordenadas gráfico 31, después del giro
(0,0)	(-1,0)	$(-\frac{\sqrt{2}}{2},\frac{\sqrt{2}}{2})$
(0,1)	(-1,1)	$(0,\sqrt{2})$
(1,0)	(0,0)	(0,0)
(1,1)	(0,1)	$(\frac{\sqrt{2}}{2},\frac{\sqrt{2}}{2})$
(x, y)	(x-1, y)	$(\frac{(x-1+y)\sqrt{2}}{2},\frac{(y+1-x)\sqrt{2}}{2})$

[Gráfico 32]



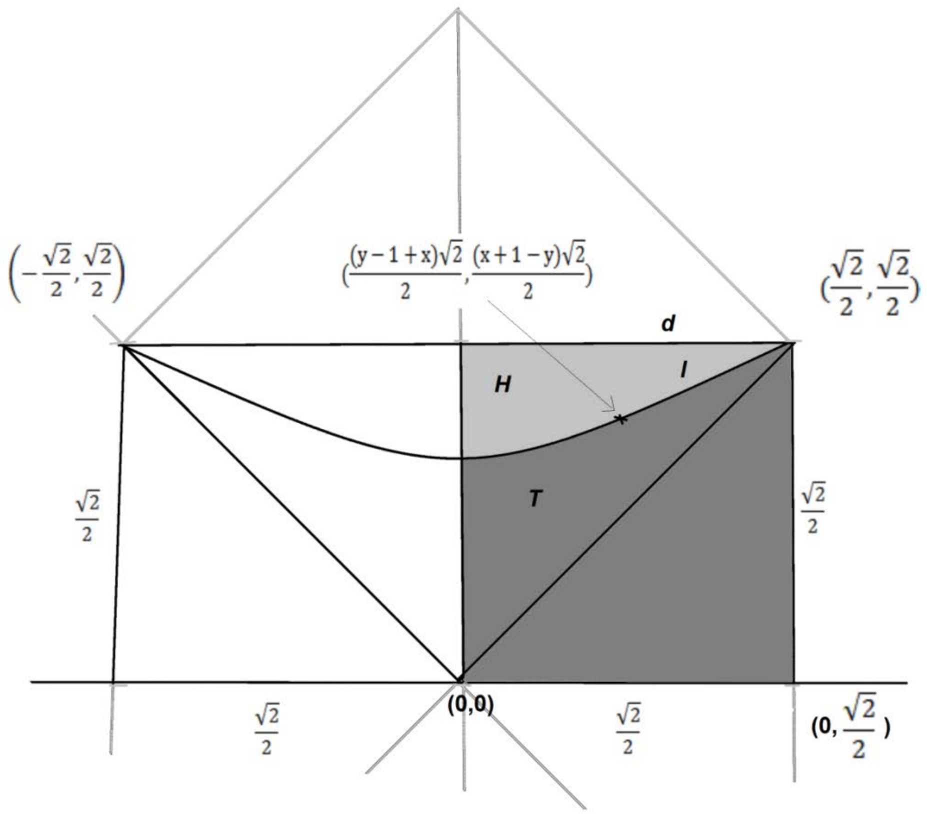
De este modo, si tenemos únicamente dos valores (x,y), podemos cargar en una hoja de cálculo los tres pares de valores de la tabla 31.

[Tabla 31]

x	y
$-\frac{\sqrt{2}}{2},$	$\frac{\sqrt{2}}{2}$
$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{2}}{2}$
$\frac{(x-1+y)\sqrt{2}}{2}$	$\frac{(y+1-x)\sqrt{2}}{2}$

Y, a continuación realizar el gráfico de dispersión correspondiente y pedirle al programa que nos proporcione la función de la curva l. Después podemos simplificar el cálculo de G considerando la mitad de su superficie (gráfico 33).

[Gráfico 33]



Necesitamos calcular la integral definida entre 0 y $\frac{\sqrt{2}}{2}$, a la que denominaremos T. Si la función de la curva es:

$$y = ax^2 + bx + c$$

Entonces podemos enunciar la fórmula:

$$T = a\frac{\sqrt{2}}{12} + b\frac{1}{4} + c\frac{\sqrt{2}}{2}$$

Si observamos el gráfico se cumple que:

$$\frac{S}{2} = T - \frac{1}{4}$$

Y, por lo tanto, multiplicando ambos términos por 4:

$$2S = 4T - 1$$

Y como:

$$IG = 1 - 2S = 1 - 4T + 1 = 2 - 4T$$

Y por lo tanto, sustituyendo el valor T que aparece en una fórmula anterior, obtenemos:

$$IG = 2 - \frac{\sqrt{2}}{3}a - b - 2\sqrt{2}c$$

Ahora ya podemos volver a plantear la cuestión inicial. Con los datos siguientes:

[Tabla 32]

	Nuevos estudiantes en el curso de referencia respecto del total de personas matriculadas (TR)	Estudiantes egresados de los nuevos en el curso de referencia (TI)
Enfermería	30,1%	70,92%
Derecho	20,3%	20,0%

La primera columna corresponde a lo que se denomina Tasa de Renovación (TR), la segunda a la Tasa de Idoneidad (TI). ¿Cuál es el porcentaje de estudiantado egresado en el curso de referencia más cuatro años respecto del conjunto de matrícula en el curso de referencia? A esta variable la denominaremos Tasa de Egresividad (TE). Se cumple que $TE = (TR)(TI)$. Por tanto, el problema se reduce a comparar los valores de TR (x) con TE (y) en dos ejes cartesianos, obtener las nuevas coordenadas (según la tabla 33) y en la parábola resultante establecer la ecuación de la curva y calcular el Índice de Gini correspondiente. Estos resultados se muestran en la tabla 33.

[Tabla 33]

	TE	Ecuación de la curva	IG
Enfermería	21,3469%	y = 0,162x2 + 5E-15x + 0,6261	0,07637
Derecho	4,0600%	y = 0,5368x2 + 2E-14x + 0,4387	0,25329

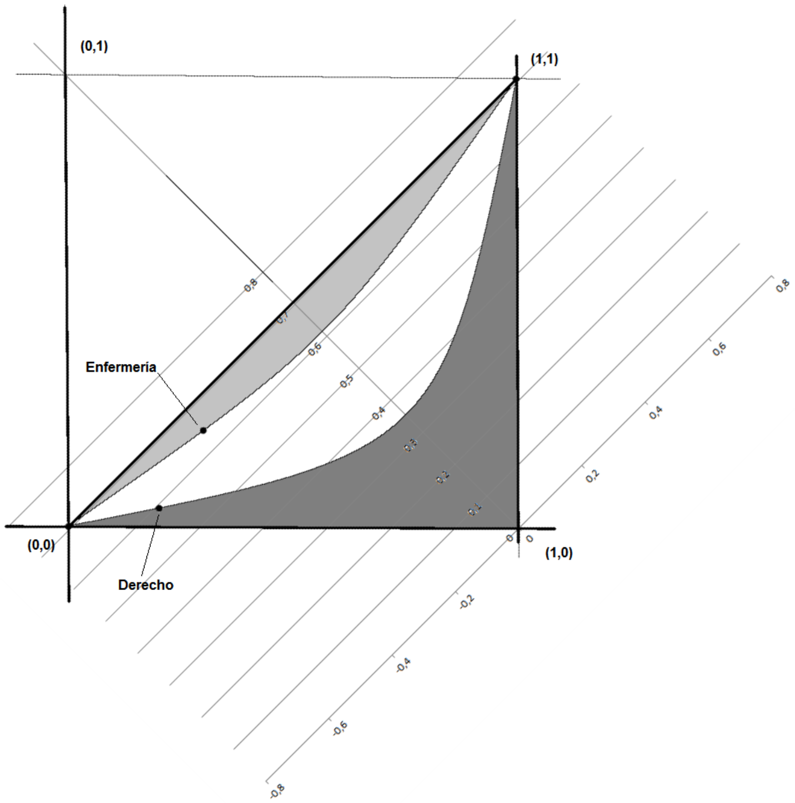
Entonces podemos hacer una comparativa entre dos titulaciones. Podemos suponer, por ejemplo, que la falta de «fluidez» e en Derecho con respecto a Enfermería es ...

$$\frac{IG_{Derecho}}{IG_{Enfermería}} = \frac{0{,}25329}{0{,}07637} = 3{,}316$$

... 3,31 veces mayor.

Podemos representar esto mismo en el gráfico 34.

[Gráfico 34]

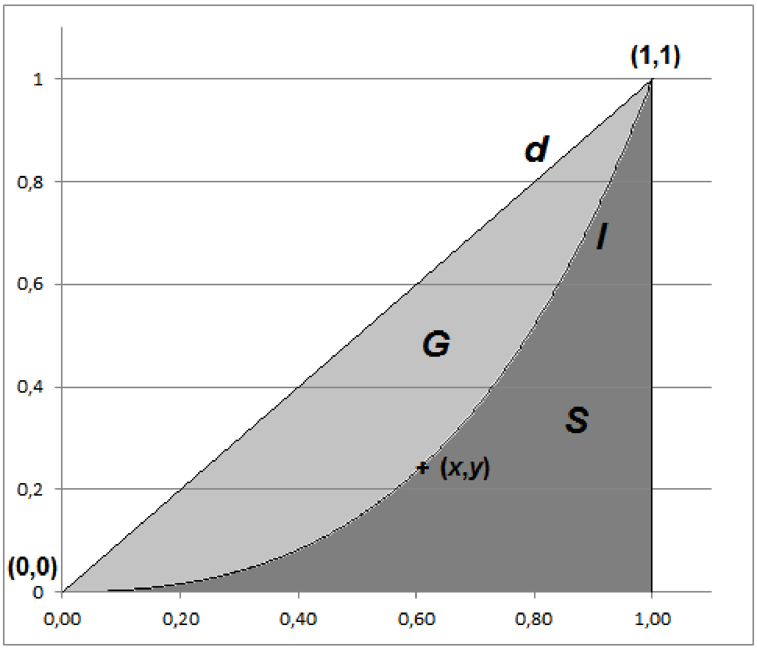


Se ha señalado en gris claro el área correspondiente al Índice de Gini (lo que antes denominábamos G) en el caso de la titulación de Enfermería. En el caso de Derecho sería el área gris clara más el área blanca. Por lo que se puede apreciar a simple vista que su índice será mayor (y por tanto, la falta de «fluidez»).

26.- ...O CON UNA CURVA EXPONENCIAL

En el epígrafe anterior utilizamos una curva parabólica, que era simètrica respecto a la línea que une los puntos (0,1) y (1,0), que es la otra diagonal, perpendicular a la bisectriz que denominamos d, para calcular el Índice de Gini (IG). Pero también podemos realizar el mismo cálculo con una curva exponencial, como la del gráfico 35.

[Gráfico 35]



(Obsérvese que la curva l no es simétrica respecto a una línea imaginaria perpendicular a d).

Podemos utilizar una función como

$$y = (2^x - 1)^a$$

Donde:

$$a = \frac{\log y}{\log(2^x - 1)}$$

Dado un par de valores (x_i, y_i), se calcula a, con la fórmula anterior, y después se puede construir una curva, asignando valores a la función general. Llevamos esos valores a una hoja de cálculo, averiguamos su línea de tendencia polinómica y después, con el procedimiento ya expuesto del cálculo de integrales, establecemos el área bajo la curva y el valor del índice.

En este caso la fundamentación se desarrolla en el **ANEXO II**. En nuestro ejemplo:

[Tabla 34]

	Nuevos estudiantes en el curso de referencia respecto del total de personas matriculadas (TR)	Estudiantes egresados de los nuevos en el curso de referencia (TI)
Enfermería	30,1%	70,92%
Derecho	20,3%	20,0%

[Tabla 35]

	x	Y [(TR)(TI)]
Enfermería	0,301	0,2134
Derecho	0,203	0,0406

Primero calculamos:

$$a_{enfermería} = \frac{\log 0{,}2134}{\log(2^{0{,}301} - 1)} = 1{,}0571$$

$$a_{derecho} = \frac{\log 0{,}0406}{\log(2^{0{,}203} - 1)} = 1{,}4820$$

Las ecuaciones resultantes (que nos proporciona la hoja de cálculo) son:

[Tabla 36]

enfermería	$y = 0{,}2295x^6 - 0{,}7915x^5 + 1{,}1124x^4 - 0{,}7309x^3 + 0{,}6125x^2 + 0{,}5687x - 0{,}0006$
derecho	$y = 0{,}3702x^6 - 1{,}2901x^5 + 1{,}9092x^4 - 1{,}2677x^3 + 1{,}1777x^2 + 0{,}1015x - 0{,}0007$

La superficie bajo curva y los Índices de Gini correspondientes son:

[Tabla 37]

	Superficie	*IG*
enfermería	0,4285	0,1429
derecho	0,3454	0,3092

$$\frac{IG_{Derecho}}{IG_{Enfermería}} = \frac{0,3092}{0,1429} = 2,163$$

De nuevo, transitamos el puente de la geometría a la aritmética, desde las formas hasta los números. Por este camino, como por un paseo vespertino, nos encontraremos con los pitagóricos, con Descartes y con aquellas personas que imaginaron matemáticamente el mundo.

27.- CONCLUYENDO

Muchas cosas quedan en el tintero. Por ejemplo, la aplicación de la teoría de las catástrofes al ámbito de la investigación educativa o el uso de espacios con más de dos dimensiones. Esto será tratado en otro texto o en la reedición de este. No voy a incluir un capítulo de bibliografía, pero sí reconocer lo inspiradoras que me resultaron las *Lecciones de matemáticas* de V. Boss, orientadas a la ingeniería, o los trabajos de A. K. Guts, Y. V. Frolova y L. A. Páutova, matematizando el orden social. Sin duda uno de los ejemplos más estimulantes está en la excelente obra del matemático L. S. Pontriaguin. Por ejemplo los desplazamientos de ejes cartesianos comentados anteriormente son herederos de su magisterio. Leer los volúmenes matemáticos de Pontriaguin no resulta fácil. Y lo auténticamente sorprendente es saber que su autor quedó ciego a los 14 años. Ciertamente, lo esencial es invisible a los ojos, como decía El Principito.

ANEXO I. Tabla de la Distribución X^2. p = Probabilidad de encontrar un valor mayor o igual que el X^2 tabulado, v= grados de libertad

v/p	0,001	0,0025	0,005	0,01	0,025	0,05	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,45	0,5
1	10,8274	9,1404	7,8794	6,6349	5,0239	3,8415	2,7055	2,0722	1,6424	1,3233	1,0742	0,8735	0,7083	0,5707	0,4549
2	13,8150	11,9827	10,5965	9,2104	7,3778	5,9915	4,6052	3,7942	3,2189	2,7726	2,4079	2,0996	1,8326	1,5970	1,3863
3	16,2660	14,3202	12,8381	11,3449	9,3484	7,8147	6,2514	5,3170	4,6416	4,1083	3,6649	3,2831	2,9462	2,6430	2,3660
4	18,4662	16,4238	14,8602	13,2767	11,1433	9,4877	7,7794	6,7449	5,9886	5,3853	4,8784	4,4377	4,0446	3,6871	3,3567
5	20,5147	18,3854	16,7496	15,0863	12,8325	11,0705	9,2363	8,1152	7,2893	6,6257	6,0644	5,5731	5,1319	4,7278	4,3515
6	22,4575	20,2491	18,5475	16,8119	14,4494	12,5916	10,6446	9,4461	8,5581	7,8408	7,2311	6,6948	6,2108	5,7652	5,3481
7	24,3213	22,0402	20,2777	18,4753	16,0128	14,0671	12,0170	10,7479	9,8032	9,0371	8,3834	7,8061	7,2832	6,8000	6,3458
8	26,1239	23,7742	21,9549	20,0902	17,5345	15,5073	13,3616	12,0271	11,0301	10,2189	9,5245	8,9094	8,3505	7,8325	7,3441
9	27,8767	25,4625	23,5893	21,6660	19,0228	16,9190	14,6837	13,2880	12,2421	11,3887	10,6564	10,0060	9,4136	8,8632	8,3428
10	29,5879	27,1119	25,1881	23,2093	20,4832	18,3070	15,9872	14,5339	13,4420	12,5489	11,7807	11,0971	10,4732	9,8922	9,3418
111	31,2635	28,7291	26,7569	24,7250	21,9200	19,6752	17,2750	15,7671	14,6314	13,7007	12,8987	12,1836	11,5298	10,9199	10,3410
12	32,9092	30,3182	28,2997	26,2170	23,3367	21,0261	18,5493	16,9893	15,8120	14,8454	14,0111	13,2661	12,5838	11,9463	11,3403
13	34,5274	31,8830	29,8193	27,6882	24,7356	22,3620	19,8119	18,2020	16,9848	15,9839	15,1187	14,3451	13,6356	12,9717	12,3398
14	36,1239	33,4262	31,3194	29,1412	26,1189	23,6848	21,0641	19,4062	18,1508	17,1169	16,2221	15,4209	14,6853	13,9961	13,3393
15	37,6978	34,9494	32,8015	30,5780	27,4884	24,9958	22,3071	20,6030	19,3107	18,2451	17,3217	16,4940	15,7332	15,0197	14,3389
16	39,2518	36,4555	34,2671	31,9999	28,8453	26,2962	23,5418	21,7931	20,4651	19,3689	18,4179	17,5646	16,7795	16,0425	15,3385
17	40,7911	37,9462	35,7184	33,4087	30,1910	27,5871	24,7690	22,9770	21,6146	20,4887	19,5110	18,6330	17,8244	17,0646	16,3382
18	42,3119	39,4220	37,1564	34,8052	31,5264	28,8693	25,9894	24,1555	22,7595	21,6049	20,6014	19,6993	18,8679	18,0860	17,3379
19	43,8194	40,8847	38,5821	36,1908	32,8523	30,1435	27,2036	25,3289	23,9004	22,7178	21,6891	20,7638	19,9102	19,1069	18,3376
20	45,3142	42,3358	39,9969	37,5663	34,1696	31,4104	28,4120	26,4976	25,0375	23,8277	22,7745	21,8265	20,9514	20,1272	19,3374
21	46,7963	43,7749	41,4009	38,9322	35,4789	32,6706	29,6151	27,6620	26,1711	24,9348	23,8578	22,8876	21,9915	21,1470	20,3372
22	48,2676	45,2041	42,7957	40,2894	36,7807	33,9245	30,8133	28,8224	27,3015	26,0393	24,9390	23,9473	23,0307	22,1663	21,3370
23	49,7276	46,6231	44,1814	41,6383	38,0756	35,1725	32,0069	29,9792	28,4288	27,1413	26,0184	25,0055	24,0689	23,1852	22,3369
24	51,1790	48,0336	45,5584	42,9798	39,3641	36,4150	33,1962	31,1325	29,5533	28,2412	27,0960	26,0625	25,1064	24,2037	23,3367
25	52,6187	49,4351	46,9280	44,3140	40,6465	37,6525	34,3816	32,2825	30,6752	29,3388	28,1719	27,1183	26,1430	25,2218	24,3366
26	54,0511	50,8291	48,2898	45,6416	41,9231	38,8851	35,5632	33,4295	31,7946	30,4346	29,2463	28,1730	27,1789	26,2395	25,3365
27	55,4751	52,2152	49,6450	46,9628	43,1945	40,1133	36,7412	34,5736	32,9117	31,5284	30,3193	29,2266	28,2141	27,2569	26,3363
28	56,8918	53,5939	50,9936	48,2782	44,4608	41,3372	37,9159	35,7150	34,0266	32,6205	31,3909	30,2791	29,2486	28,2740	27,3362
29	58,3006	54,9662	52,3355	49,5878	45,7223	42,5569	39,0875	36,8538	35,1394	33,7109	32,4612	31,3308	30,2825	29,2908	28,3361

v/p	0,001	0,0025	0,005	0,01	0,025	0,05	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,45	0,5
30	59,7022	56,3325	53,6719	50,8922	46,9792	43,7730	40,2560	37,9902	36,2502	34,7997	33,5302	32,3815	31,3159	30,3073	29,3360
31	61,0980	57,6921	55,0025	52,1914	48,2319	44,9853	41,4217	39,1244	37,3591	35,8871	34,5981	33,4314	32,3486	31,3235	30,3359
32	62,4873	59,0461	56,3280	53,4857	49,4804	46,1942	42,5847	40,2563	38,4663	36,9730	35,6649	34,4804	33,3809	32,3394	31,3359
33	63,8694	60,3953	57,6483	54,7754	50,7251	47,3999	43,7452	41,3861	39,5718	38,0575	36,7307	35,5287	34,4126	33,3551	32,3358
34	65,2471	61,7382	58,9637	56,0609	51,9660	48,6024	44,9032	42,5140	40,6756	39,1408	37,7954	36,5763	35,4438	34,3706	33,3357
35	66,6192	63,0760	60,2746	57,3420	53,2033	49,8018	46,0588	43,6399	41,7780	40,2228	38,8591	37,6231	36,4746	35,3858	34,3356
36	67,9850	64,4097	61,5811	58,6192	54,4373	50,9985	47,2122	44,7641	42,8788	41,3036	39,9220	38,6693	37,5049	36,4008	35,3356
37	69,3476	65,7384	62,8832	59,8926	55,6680	52,1923	48,3634	45,8864	43,9782	42,3833	40,9839	39,7148	38,5348	37,4156	36,3355
38	70,7039	67,0628	64,1812	61,1620	56,8955	53,3835	49,5126	47,0072	45,0763	43,4619	42,0450	40,7597	39,5643	38,4302	37,3354
39	72,0550	68,3830	65,4753	62,4281	58,1201	54,5722	50,6598	48,1263	46,1730	44,5395	43,1053	41,8040	40,5935	39,4446	38,3354
40	73,4029	69,6987	66,7660	63,6908	59,3417	55,7585	51,8050	49,2438	47,2685	45,6160	44,1649	42,8477	41,6222	40,4589	39,3353
45	80,0776	76,2229	73,1660	69,9569	65,4101	61,6562	57,5053	54,8105	52,7288	50,9849	49,4517	48,0584	46,7607	45,5274	44,3351
50	86,6603	82,6637	79,4898	76,1538	71,4202	67,5048	63,1671	60,3460	58,1638	56,3336	54,7228	53,2576	51,8916	50,5923	49,3349
55	93,1671	89,0344	85,7491	82,2920	77,3804	73,3115	68,7962	65,8550	63,5772	61,6650	59,9804	58,4469	57,0160	55,6539	54,3348
60	99,6078	95,3443	91,9518	88,3794	83,2977	79,0820	74,3970	71,3411	68,9721	66,9815	65,2265	63,6277	62,1348	60,7128	59,3347
70	112,3167	107,8079	104,2148	100,4251	95,0231	90,5313	85,5270	82,2553	79,7147	77,5766	75,6893	73,9677	72,3583	70,8236	69,3345
80	124,8389	120,1018	116,3209	112,3288	106,6285	101,8795	96,5782	93,1058	90,4053	88,1303	86,1197	84,2840	82,5663	80,9266	79,3343
90	137,2082	132,2554	128,2987	124,1162	118,1359	113,1452	107,5650	103,9040	101,0537	98,6499	96,5238	94,5809	92,7614	91,0234	89,3342
100	149,4488	144,2925	140,1697	135,8069	129,5613	124,3421	118,4980	114,6588	111,6667	109,1412	106,9058	104,8615	102,9459	101,1149	99,3341
120	173,6184	168,0814	163,6485	158,9500	152,2113	146,5673	140,2326	136,0620	132,8063	130,0546	127,6159	125,3833	123,2890	121,2850	119,3340
140	197,4498	191,5653	186,8465	181,8405	174,6478	168,6130	161,8270	157,3517	153,8537	150,8941	148,2686	145,8629	143,6043	141,4413	139,3339
160	221,0197	214,8081	209,8238	204,5300	196,9152	190,5164	183,3106	178,5517	174,8283	171,6752	168,8759	166,3092	163,8977	161,5868	159,3338
180	244,3723	237,8548	232,6198	227,0563	219,0442	212,3039	204,7036	199,6786	195,7434	192,4086	189,4462	186,7282	184,1732	181,7234	179,3338
200	267,5388	260,7350	255,2638	249,4452	241,0578	233,9942	226,0210	220,7441	216,6088	213,1022	209,9854	207,1244	204,4337	201,8526	199,3337
250	324,8306	317,3609	311,3460	304,9393	295,6885	287,8815	279,0504	273,1944	268,5987	264,6970	261,2253	258,0355	255,0327	252,1497	249,3337
300	381,4239	373,3509	366,8439	359,9064	349,8745	341,3951	331,7885	325,4090	320,3971	316,1383	312,3460	308,8589	305,5741	302,4182	299,3336
500	603,4458	593,3580	585,2060	576,4931	563,8514	553,1269	540,9303	532,8028	526,4014	520,9505	516,0874	511,6081	507,3816	503,3147	499,3335
600	712,7726	701,8322	692,9809	683,5155	669,7690	658,0936	644,8004	635,9329	628,8157	622,9876	617,6713	612,7718	608,1468	603,6942	599,3335

v/p	0,55	0,6	0,65	0,7	0,75	0,8	0,85	0,9	0,95	0,975	0,99	0,995	0,9975	0,999
1	0,3573	0,2750	0,2059	0,1485	0,1015	0,0642	0,0358	0,0158	0,0039	0,0010	0,0002	0,0000	0,0000	0,0000
2	1,1957	1,0217	0,8616	0,7133	0,5754	0,4463	0,3250	0,2107	0,1026	0,0506	0,0201	0,0100	0,0050	0,0020
3	2,1095	1,8692	1,6416	1,4237	1,2125	1,0052	0,7978	0,5844	0,3518	0,2158	0,1148	0,0717	0,0449	0,0243
4	3,0469	2,7528	2,4701	2,1947	1,9226	1,6488	1,3665	1,0636	0,7107	0,4844	0,2971	0,2070	0,1449	0,0908
5	3,9959	3,6555	3,3251	2,9999	2,6746	2,3425	1,9938	1,6103	1,1455	0,8312	0,5543	0,4118	0,3075	0,2102
6	4,9519	4,5702	4,1973	3,8276	3,4546	3,0701	2,6613	2,2041	1,6354	1,2373	0,8721	0,6757	0,5266	0,3810
7	5,9125	5,4932	5,0816	4,6713	4,2549	3,8223	3,3583	2,8331	2,1673	1,6899	1,2390	0,9893	0,7945	0,5985
8	6,8766	6,4226	5,9753	5,5274	5,0706	4,5936	4,0782	3,4895	2,7326	2,1797	1,6465	1,3444	1,1042	0,8571
9	7,8434	7,3570	6,8763	6,3933	5,8988	5,3801	4,8165	4,1682	3,3251	2,7004	2,0879	1,7349	1,4501	1,1519
10	8,8124	8,2955	7,7832	7,2672	6,7372	6,1791	5,5701	4,8652	3,9403	3,2470	2,5582	2,1558	1,8274	1,4787
11	9,7831	9,2373	8,6952	8,1479	7,5841	6,9887	6,3364	5,5778	4,5748	3,8157	3,0535	2,6032	2,2321	1,8338
12	10,7553	10,1820	9,6115	9,0343	8,4384	7,8073	7,1138	6,3038	5,2260	4,4038	3,5706	3,0738	2,6612	2,2141
13	11,7288	11,1291	10,5315	9,9257	9,2991	8,6339	7,9008	7,0415	5,8919	5,0087	4,1069	3,5650	3,1118	2,6172
14	12,7034	12,0785	11,4548	10,8215	10,1653	9,4673	8,6963	7,7895	6,5706	5,6287	4,6604	4,0747	3,5820	3,0407
15	13,6790	13,0298	12,3809	11,7212	11,0365	10,3070	9,4993	8,5468	7,2609	6,2621	5,2294	4,6009	4,0697	3,4825
16	14,6555	13,9827	13,3096	12,6243	11,9122	11,1521	10,3090	9,3122	7,9616	6,9077	5,8122	5,1422	4,5734	3,9417
17	15,6328	14,9373	14,2406	13,5307	12,7919	12,0023	11,1249	10,0852	8,6718	7,5642	6,4077	5,6973	5,0916	4,4162
18	16,6108	15,8932	15,1738	14,4399	13,6753	12,8570	11,9462	10,8649	9,3904	8,2307	7,0149	6,2648	5,6234	4,9048
19	17,5894	16,8504	16,1089	15,3517	14,5620	13,7158	12,7727	11,6509	10,1170	8,9065	7,6327	6,8439	6,1673	5,4067
20	18,5687	17,8088	17,0458	16,2659	15,4518	14,5784	13,6039	12,4426	10,8508	9,5908	8,2604	7,4338	6,7228	5,9210
21	19,5485	18,7683	17,9843	17,1823	16,3444	15,4446	14,4393	13,2396	11,5913	10,2829	8,8972	8,0336	7,2889	6,4467
22	20,5288	19,7288	18,9243	18,1007	17,2396	16,3140	15,2787	14,0415	12,3380	10,9823	9,5425	8,6427	7,8648	6,9829
23	21,5095	20,6902	19,8657	19,0211	18,1373	17,1865	16,1219	14,8480	13,0905	11,6885	10,1957	9,2604	8,4503	7,5291
24	22,4908	21,6525	20,8084	19,9432	19,0373	18,0618	16,9686	15,6587	13,8484	12,4011	10,8563	9,8862	9,0441	8,0847
25	23,4724	22,6156	21,7524	20,8670	19,9393	18,9397	17,8184	16,4734	14,6114	13,1197	11,5240	10,5196	9,6462	8,6494
26	24,4544	23,5794	22,6975	21,7924	20,8434	19,8202	18,6714	17,2919	15,3792	13,8439	12,1982	11,1602	10,2561	9,2222
27	25,4367	24,5440	23,6437	22,7192	21,7494	20,7030	19,5272	18,1139	16,1514	14,5734	12,8785	11,8077	10,8733	9,8029
28	26,4195	25,5092	24,5909	23,6475	22,6572	21,5880	20,3857	18,9392	16,9279	15,3079	13,5647	12,4613	11,4973	10,3907
29	27,4025	26,4751	25,5391	24,5770	23,5666	22,4751	21,2468	19,7677	17,7084	16,0471	14,2564	13,1211	12,1278	10,9861

v/p	0,55	0,6	0,65	0,7	0,75	0,8	0,85	0,9	0,95	0,975	0,99	0,995	0,9975	0,999
30	28,3858	27,4416	26,4881	25,5078	24,4776	23,3641	22,1103	20,5992	18,4927	16,7908	14,9535	13,7867	12,7646	11,5876
31	29,3694	28,4087	27,4381	26,4397	25,3901	24,2551	22,9762	21,4336	19,2806	17,5387	15,6555	14,4577	13,4073	12,1961
32	30,3533	29,3763	28,3889	27,3728	26,3041	25,1478	23,8442	22,2706	20,0719	18,2908	16,3622	15,1340	14,0555	12,8104
33	31,3375	30,3444	29,3405	28,3069	27,2194	26,0422	24,7143	23,1102	20,8665	19,0467	17,0735	15,8152	14,7092	13,4312
34	32,3219	31,3130	30,2928	29,2421	28,1361	26,9383	25,5864	23,9522	21,6643	19,8062	17,7891	16,5013	15,3679	14,0568
35	33,3065	32,2821	31,2458	30,1782	29,0540	27,8359	26,4604	24,7966	22,4650	20,5694	18,5089	17,1917	16,0315	14,6881
36	34,2913	33,2517	32,1995	31,1152	29,9730	28,7350	27,3363	25,6433	23,2686	21,3359	19,2326	17,8868	16,7000	15,3243
37	35,2764	34,2216	33,1539	32,0532	30,8933	29,6355	28,2138	26,4921	24,0749	22,1056	19,9603	18,5859	17,3730	15,9652
38	36,2617	35,1920	34,1089	32,9919	31,8146	30,5373	29,0931	27,3430	24,8839	22,8785	20,6914	19,2888	18,0501	16,6109
39	37,2472	36,1628	35,0645	33,9315	32,7369	31,4405	29,9739	28,1958	25,6954	23,6543	21,4261	19,9958	18,7318	17,2612
40	38,2328	37,1340	36,0207	34,8719	33,6603	32,3449	30,8563	29,0505	26,5093	24,4331	22,1642	20,7066	19,4171	17,9166
45	43,1638	41,9950	40,8095	39,5847	38,2910	36,8844	35,2895	33,3504	30,6123	28,3662	25,9012	24,3110	22,8994	21,2509
50	48,0986	46,8638	45,6100	44,3133	42,9421	41,4492	39,7539	37,6886	34,7642	32,3574	29,7067	27,9908	26,4636	24,6736
55	53,0367	51,7391	50,4204	49,0554	47,6105	46,0356	44,2448	42,0596	38,9581	36,3981	33,5705	31,7349	30,0974	28,1731
60	57,9775	56,6200	55,2394	53,8091	52,2938	50,6406	48,7587	46,4589	43,1880	40,4817	37,4848	35,5344	33,7909	31,7381
70	67,8664	66,3961	64,8990	63,3460	61,6983	59,8978	57,8443	55,3289	51,7393	48,7575	45,4417	43,2753	41,3323	39,0358
80	77,7631	76,1879	74,5825	72,9153	71,1445	69,2070	66,9938	64,2778	60,3915	57,1532	53,5400	51,1719	49,0430	46,5197
90	87,6661	85,9925	84,2854	82,5111	80,6247	78,5584	76,1954	73,2911	69,1260	65,6466	61,7540	59,1963	56,8918	54,1559
100	97,5744	95,8078	94,0046	92,1290	90,1332	87,9453	85,4406	82,3581	77,9294	74,2219	70,0650	67,3275	64,8571	61,9182
120	117,4041	115,4646	113,4825	111,4186	109,2197	106,8056	104,0374	100,6236	95,7046	91,5726	86,9233	83,8517	81,0726	77,7555
140	137,2476	135,1491	133,0028	130,7657	128,3800	125,7580	122,7476	119,0293	113,6594	109,1368	104,0343	100,6547	97,5908	93,9253
160	157,1019	154,8555	152,5564	150,1583	147,5988	144,7834	141,5475	137,5457	131,7560	126,8700	121,3457	117,6791	114,3496	110,3592
180	176,9652	174,5799	172,1373	169,5879	166,8653	163,8682	160,4206	156,1526	149,9687	144,7413	138,8205	134,8843	131,3050	127,0114
200	196,8359	194,3193	191,7409	189,0486	186,1717	183,0028	179,3550	174,8353	168,2785	162,7280	156,4321	152,2408	148,4262	143,8420
250	246,5387	243,7202	240,8297	237,8085	234,5768	231,0128	226,9048	221,8059	214,3915	208,0978	200,9387	196,1604	191,8020	186,5537
300	296,2700	293,1786	290,0062	286,6878	283,1353	279,2143	274,6901	269,0679	260,8781	253,9122	245,9727	240,6631	235,8126	229,9620
500	495,3734	491,3709	487,2569	482,9462	478,3231	473,2099	467,2962	459,9261	449,1467	439,9360	429,3874	422,3034	415,8081	407,9458
600	594,9938	590,6057	586,0930	581,3623	576,2859	570,6681	564,1661	556,0560	544,1801	534,0185	522,3654	514,5285	507,3385	498,6219

ANEXO II

Fundamentación del cálculo del Índice de Gini con curvas parabólicas

Hay diversos procedimientos para calcular *IG*. Aquí utilizaremos el que parte del cálculo de la superficie bajo curva, es decir, el área de G delimitada por la curva l y las líneas que unen los puntos (0,0) y (1,0), por un lado, y (1,0) y (1,1) por otro lado. Esta superficie es la que hemos denominado G en el diagrama 1 y presenta un color gris más oscuro.

Si fl es la función de la curva l, entonces:

$$\mathrm{S} = \int_0^1 \mathrm{fl}\ (\mathrm{x})\ \ \mathrm{dx}$$

Y, como ya se dijo que:

$$\mathrm{IG} = \ 1 - 2\mathrm{S}$$

entonces:

$$\mathrm{IG} = 1 - 2\int_0^1 \mathrm{fl}\ (\mathrm{x})\ \ \mathrm{dx}$$

Siendo fl la función polinómica de la curva l. Si por ejemplo fuera una función polinómica de grado 6, entonces:

$$S = \int_0^1 (ax^6 + bx^5 + cx^4 + dx^3 + ex^2 + fx + g)dx =$$

$$= a\int_0^1 x^6\,dx + b\int_0^1 x^5\,dx + c\int_0^1 x^4\,dx + d\int_0^1 x^3\,dx + e\int_0^1 x^2\,dx + f\int_0^1 x\,dx + g\int_0^1 dx =$$

$$= a\left[\frac{x^7}{7}\right]_0^1 + b\left[\frac{x^6}{6}\right]_0^1 + c\left[\frac{x^5}{5}\right]_0^1 + d\left[\frac{x^4}{4}\right]_0^1 + e\left[\frac{x^3}{3}\right]_0^1 + f\left[\frac{x^2}{2}\right]_0^1 + g\,[x]_0^1 =$$

$$= \frac{a}{7} + \frac{b}{6} + \frac{c}{5} + \frac{d}{4} + \frac{e}{3} + \frac{f}{2} + g$$

Y por tanto:

$$IG = 1 - 2\left(\frac{a}{7} + \frac{b}{6} + \frac{c}{5} + \frac{d}{4} + \frac{e}{3} + \frac{f}{2} + g\right)$$

Por lo que:

$$IG = 1 - \frac{2a}{7} - \frac{b}{3} - \frac{2c}{5} - \frac{d}{2} - \frac{2e}{3} - \mathrm{f} - 2g$$

Fundamentación del cálculo del Índice de Gini con curvas exponenciales

Los ejemplos producen los gráficos y las ecuaciones siguientes.

Gráfico Enfermería:

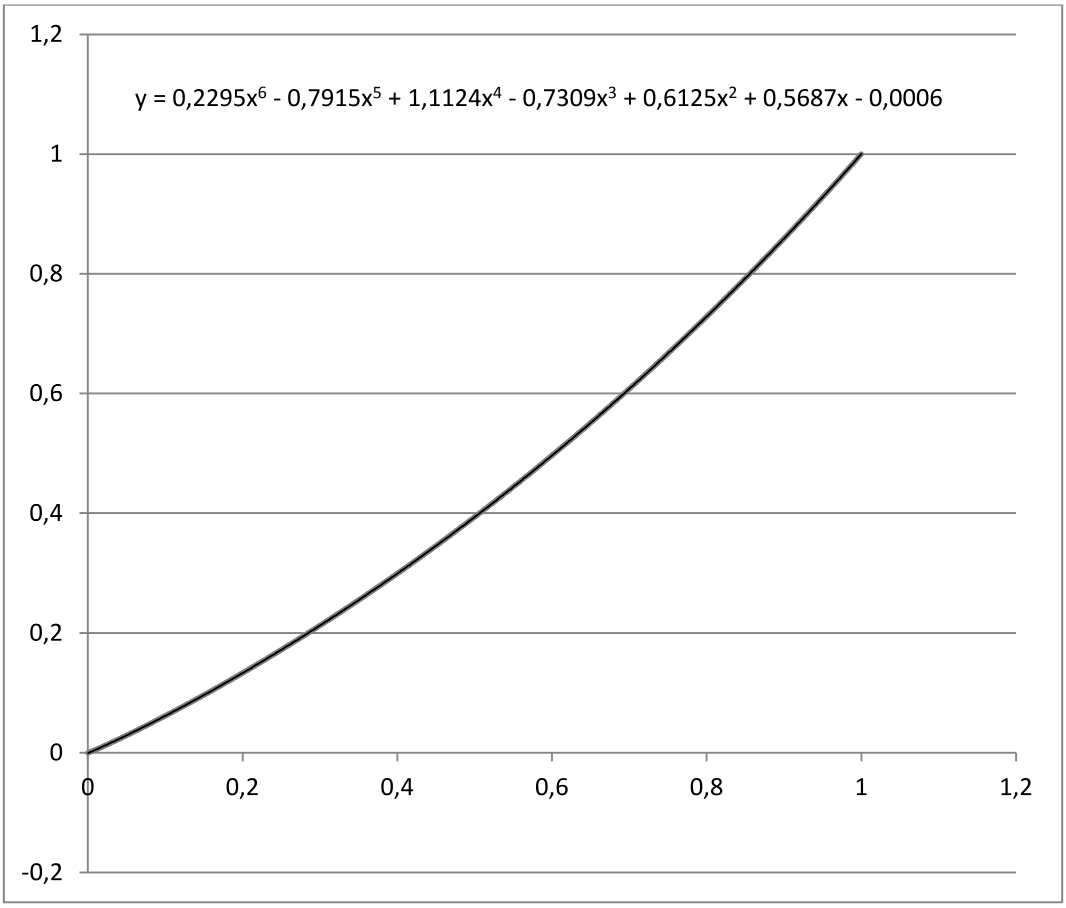
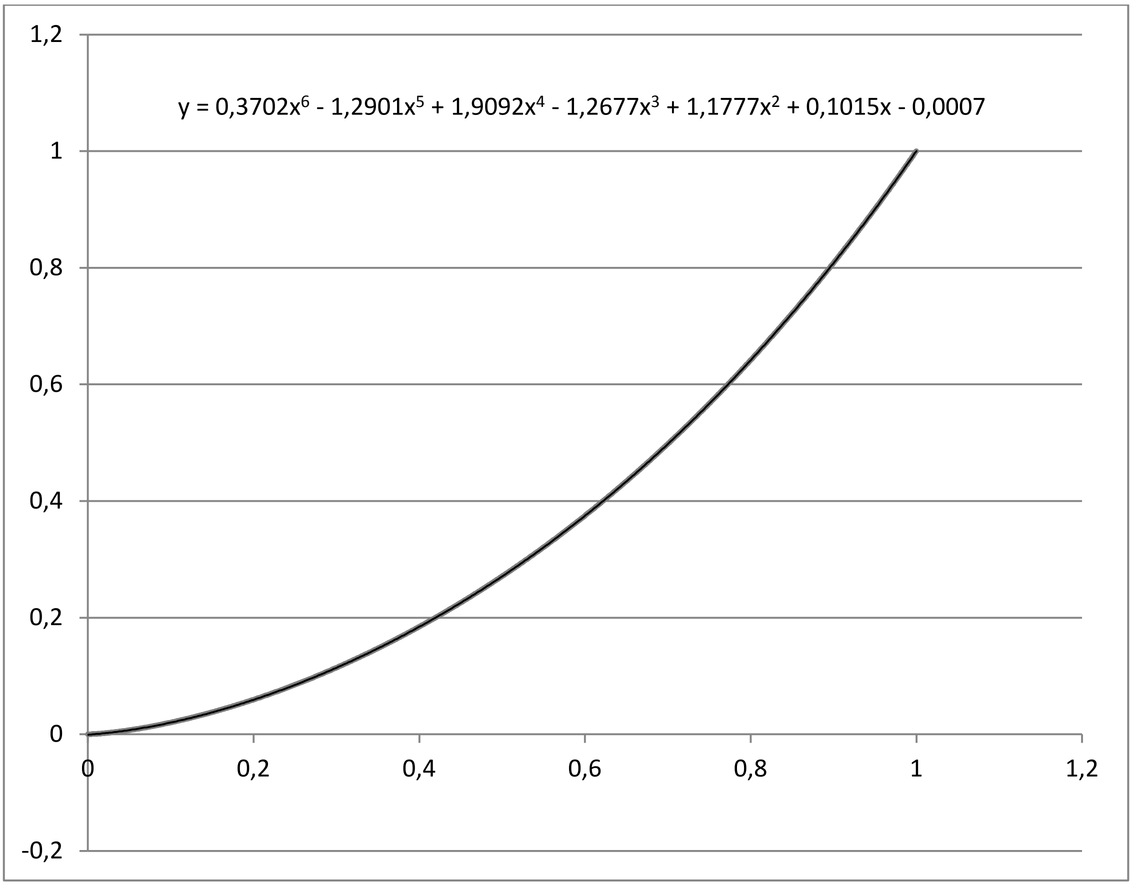


Gráfico Derecho:



Que se integran para calcular la superficie bajo curva.

ÍNDICE