

LA ESTADÍSTICA R^2 EN INFORMES PERICIALES. UN EJEMPLO DE SU APLICACIÓN.

Por **José Antonio Granados**, Economista Forense

jose.granados@grossman.es

Socio de **GROSSMAN, MINGOT, GRANADOS Y ASOCIADOS**

Socios de la Asociación de Peritos de la Comunidad de Madrid www.apajcm.com

Por **Claudia Granados**, Grado en Físicas y Máster en Ingeniería Nuclear

Colaboradora de **GROSSMAN, MINGOT, GRANADOS Y ASOCIADOS**

Primero. Introducción.

En el desarrollo del quehacer económico pericial nos hemos encontrado con la necesidad de trabajar sobre poblaciones de datos económicos en las que tenemos que pronunciarnos acerca de su bondad o no en algún aspecto particular que se nos solicita.

En unos casos solemos hacer un estudio de estas series de datos representándolas mediante gráficas con el fin de observar a alto nivel la relación o no relación que pueda existir entre ellos, y concluir manifestando que es evidente la correlación o no entre ellos. En otras ocasiones nos enfrascamos en solicitar documentación y más documentación con el objetivo de que ésta no se aporte, y poder concluir que los datos económicos son cuestionables por falta de soporte suficiente. Y así un largo etcétera.

Hace años publicamos un artículo sobre el uso de la estadística *r de Spearman* que nos permitía medir la fuerza de una relación entre observaciones por pares de dos variables cuando los datos se encuentran en forma jerarquizada.

Hoy presentamos una nueva estadística, la *R cuadrado*, su uso nos ha ayudado como Profesionales en el quehacer forense, a dictaminar de manera más objetiva sobre cifras sujetas a duda judicial y soportar en mejor medida nuestras conclusiones. Este artículo surge de aplicaciones prácticas en Informes económico-forenses.

Segundo. Concepto de R^2 .

En estadística, el coeficiente de determinación denominado R^2 y pronunciado como **R cuadrado** y escrito también como **R2**, es una variable usada en el contexto de un modelo estadístico cuyo principal propósito es predecir futuros resultados o probar una hipótesis. El coeficiente determina la calidad representativa del modelo sobre los datos a estudiar, e indica la proporción de variación total que está siendo explicada por la regresión del modelo.

Al conjunto de datos se le debe hallar de manera previa la curva de regresión, que tiene carácter de línea media y trata de resumir o sintetizar la información suministrada por los datos. La curva de regresión refleja una recta representativa de la evolución de los datos; su objetivo es ajustar la mejor curva posible representativa de los datos a estudiar, es decir, aquella en la que las distancias entre los puntos representados y los puntos correspondientes a la misma sean mínimos. Dado que tiene carácter de línea media (de promedio, en definitiva), deberá ir acompañada siempre de una medida que exprese su representatividad, es decir, de lo buena que es la curva, ya que al haber sido obtenido de una media, no tenemos garantías de que sea correcta.

La *curva de regresión* puede ser de un tipo u otro dependiendo de la distribución de la población de datos que queremos estudiar; en la herramienta Excel existe la posibilidad de ajustar la curva de regresión a una curva lineal, a una exponencial y a una logarítmica entre otras. Para una misma población de datos existirá una curva que sea más representativa que las demás, el modo de saber qué curva se ajusta mejor a nuestros datos es mediante el valor del coeficiente de determinación. Cada curva de regresión poseerá un valor para su R^2 , dicho valor se encuentra dentro de $0 < R^2 < 1$. Por tanto:

- Si $R^2 = 1$ no hay residuos: habrá una dependencia funcional. Cuanto más se acerque el valor de R^2 a la unidad, mayor poder explicativo tendrá el modelo de regresión.
- Si $R^2 = 0$ entonces la curva de regresión no explica en absoluto ninguna de las variaciones de la variable Y, de modo que o bien el modelo es inadecuado, o bien las variables son independientes unas de otras.

En definitiva, seguimos confeccionando las gráficas que exponemos en nuestros informes, como hasta ahora, a partir de la población de datos a la que le incorporamos:

- (a) la curva de regresión o tendencia –que en muchos informes ya usamos, y
- (b) la estadística *R cuadrado*.

Tercero. Cálculo de la R cuadrado.

El ajuste mediante la curva de regresión y su correspondiente valor de R^2 se puede calcular mediante la herramienta Excel. Se necesita de manera obligada el conjunto de datos a estudiar propiamente dichos y la referencia a la que están relacionados. Asimismo, debemos de calcular la curva de regresión o tendencia a la que asociar la R^2 .

A continuación, exponemos un conjunto de datos, que luego nos servirá para el ejemplo base de este artículo.

Supongamos que el pueblo de Esquina, situado al sur de la provincia de Corrientes, cobra a sus vecinos la tasa de basuras. Cada año publica en el boletín provincial la oportuna Ordenanza en la que se expone la tarifa y que suponemos es la siguiente:

Tabla 1

Categorización de la cuota tributaria					Tarifa 2009-2010 (€)	Referencia
I.	Vivienda				39,22	1
II.	Alojamientos	Hoteles y Apartahoteles	hasta 15 hab.		270,43	2
			de 16 a 30		402,18	3
			de 31 a 50		577,85	4
			de 51 en adelante		866,77	5
		Pensiones y casas de huéspedes	hasta 15 hab.		128,00	6
			de 16 a 30 hab.		196,47	7
			de 31 a 50 hab.		250,86	8
			de 51 en adelante		401,98	9
III.	Servicios	Escuelas privadas, talleres-escuela y clínicas particulares			191,97	10
		Despachos profesionales, gestorías, oficinas, academias, autoescuelas, guarderías, ect			154,38	11
		Entidades bancarias y Cajas de Ahorros			258,18	12
		Garajes		hasta 10 plazas	126,58	13
				de 11 a 30 plazas	167,59	14
				de 31 en adelante	249,66	15
IV.	Establecimientos comerciales	Alimentación	Autoservicio	supermercados	5.576,48	16
				251 m2 ≤ S ≤ 400 m2	693,41	17
				120 m2 ≤ S ≤ 250 m2	254,77	18
				Resto	202,25	19
			puestos en mercado		154,40	20
		No alimentación			141,95	21
V.	Establecimientos Industriales	Restauración	Restaurantes	S≤ 100 m2	346,71	22
				101 m2 ≤ S ≤ 250 m2	693,41	23
				251 m2 ≤ S ≤ 500 m2	924,55	24
				S≥501 m2	2.118,77	25
		Cafeterías y otros Bares y Cafés	S≤ 100m2	196,47	26	
			S≥101m2	312,04	27	
		Bares de Categoría Especial			312,04	28
		Peluquerías y Salones de Belleza			196,47	29
		Talleres de Reparación			196,47	30
		Industrias			265,81	31

Comentarios (i): m² = metros cuadrados; (ii) S = superficie; (iii) hab =habitaciones; (iv) la columna **Referencia** es un añadido de los autores de este artículo.

Vamos a calcular las curvas de dispersión y sus correspondientes R^2 utilizando Excel para dos subepígrafes de la categorización de la anterior gráfica: Alimentación y Hoteles y Apartahoteles.

Para Alimentación la nube de datos es:

Tabla 2

Categorización de la cuota tributaria			Tarifa 2009-2010
Alimentación	Autoservicio	supermercados	5.576,48
		$S \geq 400 \text{ m}^2$	693,41
		$120 \text{ m}^2 \leq S \leq 399 \text{ m}^2$	254,77
		resto	202,25
	puestos de mercado		154,40

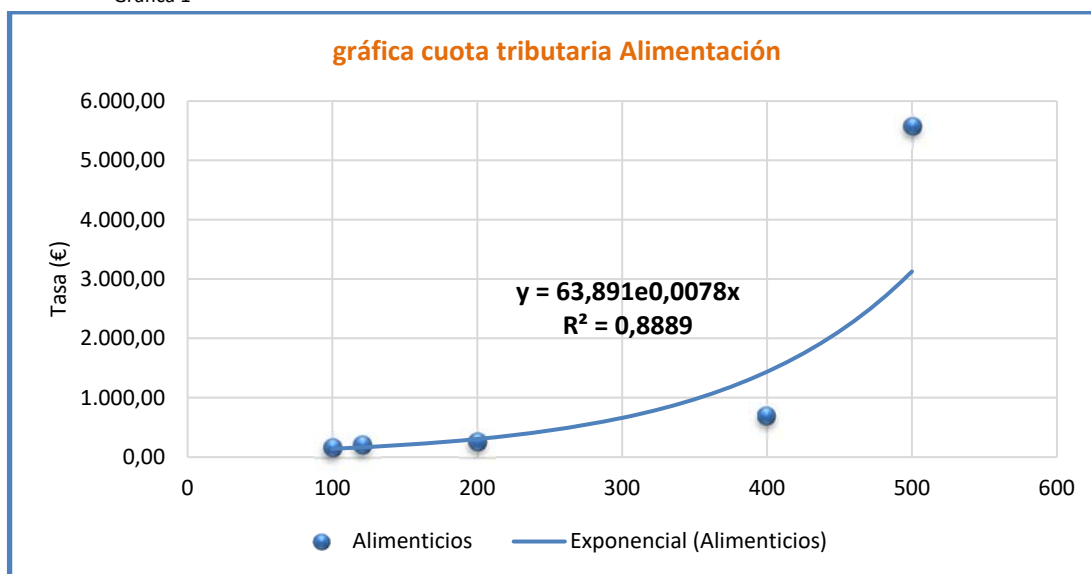
Entramos en Excel, hacemos click en *Insertar* → *gráficos (dispersión)* → *seleccionar datos* y construimos el siguiente gráfico:

- Nombre de la serie: Alimentación
- Valores X de la serie: seleccionamos con el cursor los valores de la tarifa: 5.576,48; 693,41; y sucesivos.
- Valores Y de la serie: debemos de seleccionar la columna de superficies. Dado que deben ser números, les asignamos –para este artículo- los valores 500, 400, 200, 120 y 100 (para los puestos de mercado).

A continuación, sobre cualquier punto de la gráfica que hemos creado hacemos click sobre: *Presentación* → *Línea de Tendencia* y seleccionamos las siguientes opciones: *Exponencial*, *Presentar ecuación en el gráfico* y *Presentar el valor R cuadrado en el gráfico*. De esta manera sencilla ya tenemos la información estadística con la que vamos a trabajar.

Para el ejemplo del presente artículo hemos elegido la *curva exponencial* porque es la que posee el valor de R^2 más próximo a la unidad con respecto a las demás curvas, y, además, se trata de la curva más habitual en el mundo de la economía.

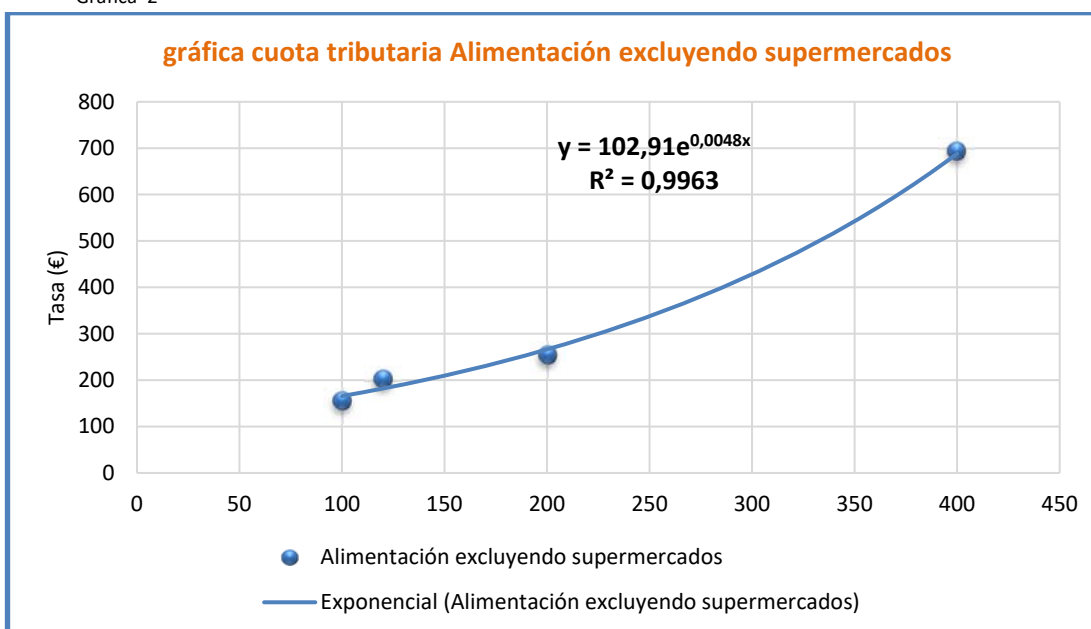
Gráfica 1



Así como se ha confeccionado esta gráfica, podemos preparar las gráficas de los otros sub epígrafes. Es sencillo.

Antes de terminar, deseamos exponer la *gráfica Alimentación* pero excluyendo el valor extremo y objeto de la controversia jurídica y de nuestro informe pericial. Rehacemos el gráfico excluyendo el valor Supermercados, y el resultado es el siguiente:

Gráfica 2



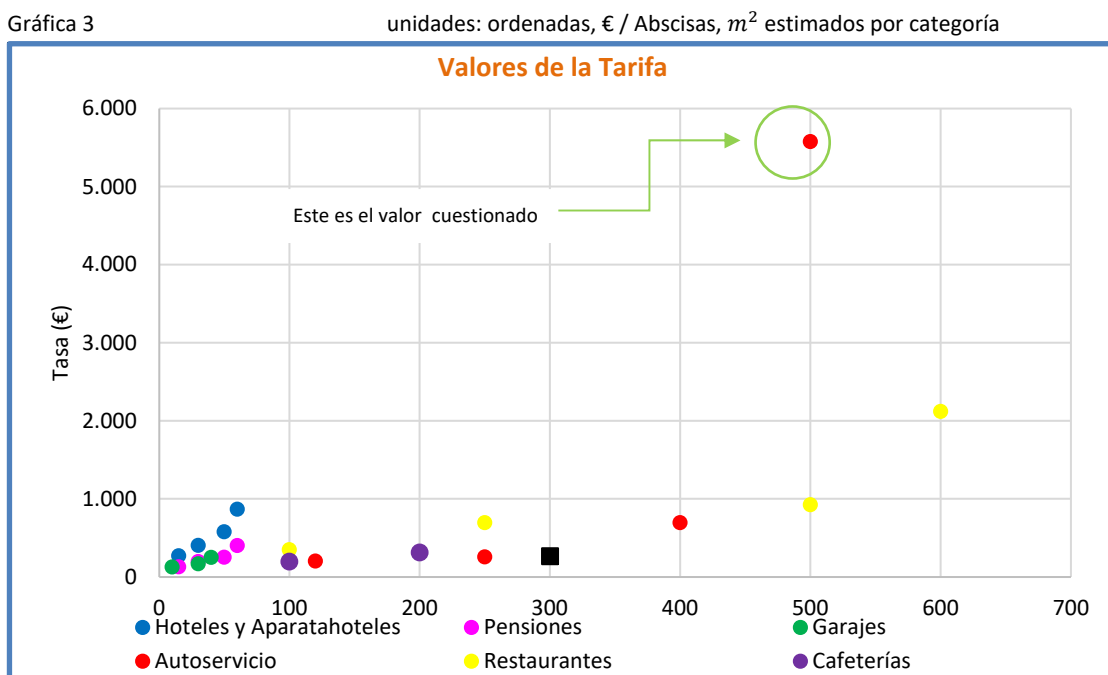
Podemos observar que la gráfica 2 posee una curva de tendencia mucho más racional en términos matemáticos que la Gráfica 1, ya sea por la forma de la curva como por el valor del coeficiente de dispersión que en la gráfica 2 es más cercano a la unidad, a

uno. A partir de este hecho podemos empezar a concluir con soporte estadístico y manifestar que el valor crítico “supermercados” está fuera de rango en términos de análisis estadístico.

Cuarto. Un ejemplo del uso de la R2 en el quehacer forense.

La controversia judicial surge porque un contribuyente que figura categorizado dentro del padrón IV, Establecimientos comerciales-Alimentación-Autoservicio-Supermercados, manifiesta que está discriminado y, por supuesto, paga muy por encima de lo que le corresponde en equidad fiscal, en el denominado principio de equivalencia que ha de regir en materia de tasas.

La gráfica de dispersión de los valores de la tarifa expuestos en la tabla 1 anteriores es la siguiente:



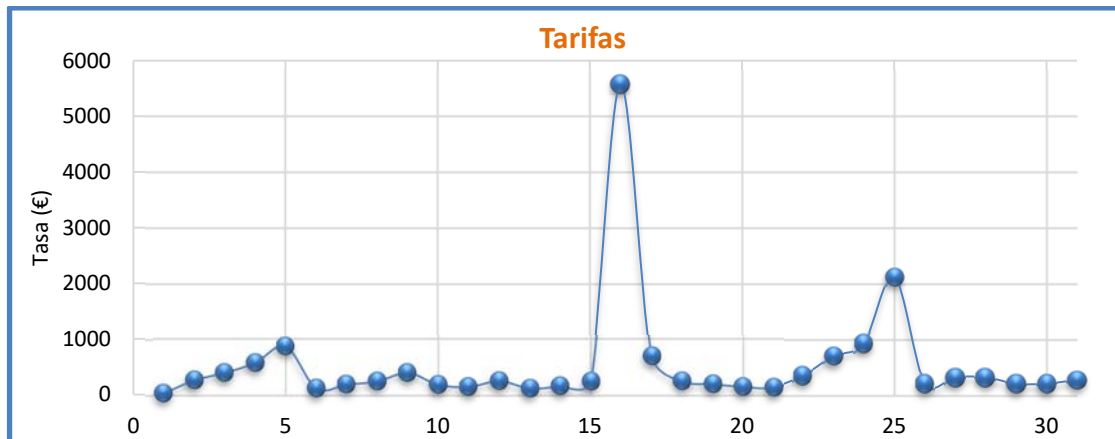
Fuente: elaboración a partir de tabla 1.

Comentario: No se expone la cuota referida al epígrafe I. Viviendas porque el valor se pierde en la gráfica al ser un valor más pequeño respecto de los otros epígrafes y estar cercano a cero.

La gráfica de los valores de todos los epígrafes de la tarifa es la siguiente:

Gráfica 4

unidades: ordenadas: € / abscisas: número correlativo de referencia



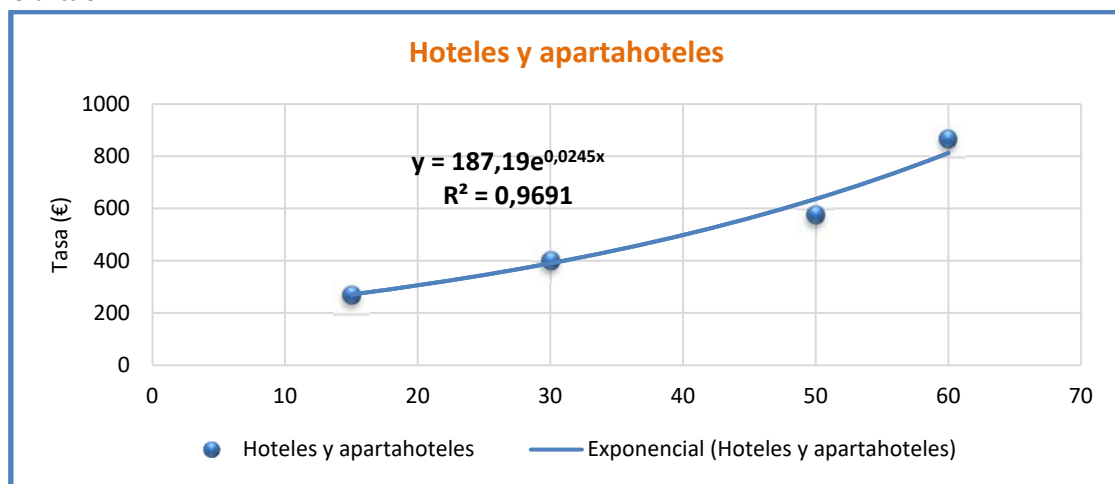
Fuente: elaboración a partir de la tabla 1.

Se observa que el valor asignado al epígrafe Supermercados por el municipio de Esquina está fuera de rango. Podríamos afirmarlo y terminar el informe manifestando que es evidente la discriminación por obvio, añadiendo la gráfica 4, y concluyendo que la cuota tributaria referida al epígrafe cuestionado está fuera de rango.

No obstante, consideramos que el trabajo profesional que soporta la conclusión sería insuficiente a nuestra opinión, aunque le añadiéramos 2 ó 3 gráficas más.

La segunda gráfica que proponemos su cálculo es el referido al sub epígrafe II, Hoteles y apartahoteles y que es el siguiente:

Gráfica 5



Se observa la racionalidad de este epígrafe al confeccionarse la tarifa y lo mismo ocurre con el resto de casi todos los sub epígrafes.

A continuación, exponemos un resumen de las R^2 de las diferentes curvas de regresión del tipo exponencial que surgirían de cada una de las categorías para la tarifa de la tasa

Tabla 3

Categorías	R^2
Hoteles y apartahoteles	0,9807
Garajes	0,9182
Restaurantes	0,8968
Pensiones	0,9482
Alimentación	0,8890
Alimentación excluyendo supermercados	0,9861

Podemos observar en la tabla 3 que la Tarifa, tomando en cuenta el valor referido a Alimentación excluyendo Supermercados, está construida por el municipio de Esquina de manera racional, porque en términos económicos por el valor de R^2 de las curva de regresión de los distintos sub epígrafes, corresponden prácticamente al 0,99, excepto en *Alimentación incluyendo supermercados* que se encuentra por debajo del 0,90.

El párrafo anterior es una conclusión con base estadística que supera a la racionalidad visual de una gráfica. Hay que destacar que en principio en las R^2 de Alimentación (0,8890 versus 0,9861) parezca una gran diferencia, la realidad en términos estadísticos si lo es.

Quinto. Valor alternativo de la tasa

Dado que tenemos la ecuación de la curva para la subcategoría Alimentación sin Supermercados, a partir de la misma se puede despejar y calcular el valor teórico de la tasa de recogida de basuras para el supermercado que denuncia la arbitrariedad de su tasa. Hemos escogido la curva Alimenticios sin Supermercado frente a Alimentación porque es la que tiene mejor valor de R^2 . El expediente judicial indica que el establecimiento denunciante tiene una superficie de 455,10 m². Los cálculos y resultado son:

Paso 1	Fórmula de cálculo	$y = 102,91e^{(0,0048x)}$
Paso 2	Sustituimos la X por 455,10 m ²	$y = 102,91 * e^{(0,0048 * 455,1)}$
Cálculo	Resultado de y La fórmula de cálculo en Excel es: =102,91*EXP(0,0048*455,1)	914,46 euros

Por lo tanto, si la Tasa impuesta a Supermercados siguiera la curva correspondiente a los demás puntos del sector Alimentación, el valor de la tasa fiscal a pagar por el contribuyente Supermercados sería igual a 914,46€. A día de hoy la Tasa impuesta por el municipio de esquina es igual a 5.577 euros, lo que implica un aumento del 600% aproximadamente sobre el valor teórico económico.

Asimismo, podemos despejar la ecuación y conocer cuál sería la superficie del establecimiento de alimentación autoservicio (el valor de X) para la tasa de 5.577 euros, aplicando la ecuación exponencial antes descrita, y el resultado es de 831,76 m² (La fórmula de cálculo en Excel es: $=\text{LN}(5576,48)-\text{LN}(102,91))/0,0048$). Es decir, la Tasa que existe e impuesta en términos fiscales actualmente a un supermercado de 455 m² equivale a una tasa para un establecimiento de aproximadamente el doble de superficie.

Sexto. Cuándo y cómo utilizar la R² en un trabajo pericial

El coeficiente de determinación R², se utilizará cuando el estudio se centre en datos correlacionados. A continuación exponemos una serie de preguntas que nos ayudarán a conocer cuando aplicar la técnica.

¿Cuáles son esos datos que solemos utilizar en la economía forense susceptibles de ser analizados con esta técnica?

Deben ser datos correlacionados, es decir, datos que conlleven una relación de dependencias entre ellos, datos como los que han sido objeto de estudio en este artículo, relacionando por ejemplo superficie con tasa. También se utiliza en informes relacionados con cártel en denuncias de la CNMC.

¿Cuándo es conveniente la aplicación de la técnica?

Cuando deseamos concluir sobre la relación existente en una población de datos para calificar sobre algún aspecto existente respecto del total.

¿Qué hacer con el resultado R²?

El valor de R² de la curva de regresión puede ser entre 1 y 0 y nos indica la relación o no existente entre ellos. Servirá de base para las conclusiones del Informe en el que estamos trabajando con soporte estadístico.

¿Qué se requiere del Perito Economista para trabajar con la R²?

Tener conocimientos básicos-medios de estadística. No es necesario ser un súper-experto porque no seríamos Economistas sino Estadísticos, pero si se requiere una vocación hacia la estadística. Hay muchos y buenos *libros de estadística* pero sin esa vocación, su lectura y aplicación en los informes se hace difícil o imposible. Un problema añadido es el desconocimiento de los operadores jurídicos de la estadística aplicada.

Séptimo. Conclusión.

Después de aplicar la estadística **R2** en nuestro quehacer económico-forense consideramos que podemos concluir lo siguiente:

1. La calidad de nuestro trabajo y Dictamen aumentó al incorporar en nuestras conclusiones profesionales una herramienta objetiva que minimiza (o enriquece) nuestra subjetividad profesional.
2. Defendemos mejor nuestro dictamen y sus conclusiones en los Actos de Ratificación al incorporar herramientas objetivas de muy difícil contestación por la Parte que se considera afectada de manera negativa.
3. Ayudamos en el proceso judicial para que el Juez pueda concluir en sus Sentencias, a través de nuestros informes y en los aspectos económico-forenses, de manera más correcta.

Este artículo es de libre acceso. Los autores agradecen, si corresponde, que se haga referencia a él en Informes o trabajos profesionales.