

Transformaciones de datos en la elaboración de estudios salariales

Albert Fornieles*

Universidad Autónoma de Barcelona, España

ARTICLE INFORMATION

Manuscrito recibido: 11/04/2013
Revisión recibida: 06/08/2013
Aceptado: 06/08/2013

Palabras clave:

Estimación salarial
Regresión
Transformación de datos
Transformación logarítmica

Keywords:

Salary estimation
Regression
Data transformation
Logarithmic transformation

RESUMEN

El presente trabajo pretende mostrar algunos aspectos de la preparación y el tratamiento de los datos procedentes de encuestas salariales con vistas a la elaboración de informes retributivos. Veremos cómo, tras probar diferentes transformaciones alternativas, se puede comprobar que mediante una transformación logarítmica es posible mejorar la asimetría y la curtosis de las variables del modelo (facturación y salario), así como convertir en lineal una relación entre dos variables que con los datos originales no lo es, lo que facilita el análisis de los datos al permitir utilizar el modelo de regresión. Estudiaremos la relación entre el salario y el tamaño de la empresa, ya que esta relación es la que mejor predice el salario de mercado que debe obtener un empleado en un puesto determinado. Presentamos un ejemplo de la aplicación del modelo de regresión para el estudio de esta relación.

© 2013 Colegio Oficial de Psicólogos de Madrid. Todos los derechos reservados.

Data transformations in salary surveys

ABSTRACT

This article intends to show some of the aspects involved when elaborating and processing data coming from salary surveys to produce retribution reports. We will see how after trying different alternative transformations it is possible to verify that, through a logarithmic transformation, the asymmetry and kurtosis of a model's variables can be improved (revenues and salary) as well as a relation between two variables converted in lineal –something that would have been impossible to achieve with the original data. This would make data analysis easier, because it allows the use of the regression model. We will study the relation between salary and company size, because this relationship is the best to predict the market salary that is earned by a specific jobholder. We give an example of application of a regression model for studying this relationship.

© 2013 Colegio Oficial de Psicólogos de Madrid. All rights reserved.

El presente trabajo pretende mostrar algunos aspectos de la preparación y el tratamiento de los datos procedentes de encuestas salariales con vistas a la elaboración de informes retributivos. Para la exposición del tema hemos escogido un ejemplo desarrollado con datos reales, en el que la utilización de las transformaciones ha permitido mejorar la posibilidad de análisis e interpretación de los datos. El tema presentado se ubica en el contexto de uno de los aspectos que están tomando mayor auge en la gestión de los recursos humanos: la política salarial y, más concretamente, los estudios salariales.

A lo largo del artículo estudiaremos la relación que se establece entre el salario y el tamaño de la empresa. Esta relación es crucial en la confección de estudios salariales, ya que el tamaño de la empresa

es la variable que mejor predice el salario de mercado –el mercado salarial se define como el colectivo de empresas que están en disposición de contratar a las mismas personas; obviamente, este mercado no es el mismo para todos los puestos de la empresa– que debe obtener un empleado que ocupa un puesto determinado.

En la figura 1 (reproducida de CEINSA, 2012) se pueden apreciar los mercados de referencia para los puestos de los distintos niveles jerárquicos y áreas funcionales. En efecto, el tamaño de la organización es la variable que muestra mayor influencia en la remuneración, especialmente en lo que concierne a los cargos ejecutivos.

Aunque existen diversos indicadores para medir el tamaño de una organización –valor de los activos, valor añadido, beneficios– los más significativos y fáciles de obtener de manera fiable son la cifra de facturación (volumen de ventas) y el número de empleados en la nómina de la empresa (plantilla).

A efectos retributivos, el volumen de ventas es un predictor más eficaz que la plantilla para la mayoría de los cargos, por lo que pres-

*La correspondencia sobre este artículo debe enviarse a Albert Fornieles Deu. Facultad de Psicología. Edificio B Campus UAB. 08193 Bellaterra. E-mail: Albert.fornieles@uab.cat

ÁREA \ NIVEL	Administración, informática y personal	Ventas y marketing	Producción y técnica
Directivos	Sector genérico Zona amplia	Sector afín Zona amplia	Sector afín Zona amplia
Jefes y mandos	Sector genérico Zona cercana	Sector afín Zona cercana	Sector muy afín Zona cercana
Personal técnico	Sector genérico Zona próxima	Sector afín Zona próxima	Sector muy afín Zona próxima

Figura 1. Mercados de referencia para los puestos de los distintos niveles jerárquicos y áreas funcionales.

cindiremos de los datos de la plantilla (v. gr., CEINSA, 1995, 2012; PE, 2006; Watson Wyatt, 2004).

Las razones teóricas que justifican la influencia del tamaño de la empresa sobre el salario son varias (v. gr., CEINSA, 1995, 2012; PE, 2006; Watson Wyatt, 2004).

- La responsabilidad y complejidad de un cargo directivo aumenta a medida que se incrementa el valor de las magnitudes económicas asociadas al puesto de trabajo: número de personas que dependen de él, presupuestos de gastos, de inversión, valor de los activos que maneja, etc. Todos estos parámetros están relacionados con el tamaño de la organización.
- Las empresas grandes suelen tener más escalones retributivos que las pequeñas, por lo que el abanico salarial –la diferencia entre los sueldos más altos y los bajos– se amplía normalmente a medida que aumenta el tamaño de la organización.
- La incidencia económica del coste de la estructura directiva suele ser proporcionalmente mayor en una empresa pequeña que en una empresa grande, por lo que ésta se halla en condiciones de soportar costes salariales *per capita* más elevados.

En este trabajo veremos cómo la relación entre las variables predictoras (facturación) y criterio (salario) no es lineal, por lo que para la realización del modelo de regresión se hace necesaria la transformación de las variables. El objetivo del artículo es mostrar que es la más adecuada para este tipo de estudios.

Formalmente, en el ámbito científico, podemos definir la transformación de una variable como el resultado de aplicar alguna expresión idéntica sobre todos sus valores, de manera que cada uno de ellos guarde una misma correspondencia con los datos de la variable original. En esta expresión pueden intervenir variables, constantes, operadores aritméticos y funciones (Domènech, 1999).

En este mismo sentido, el *Diccionari de la llengua Catalana (Institut d'Estudis Catalans)*, define transformación en su aceptación matemática como una “aplicación biyectiva entre dos subconjuntos de un espacio” (traducido por el autor). En definitiva, como señala Salvador (1996), las transformaciones consisten en obtener valores numéricos diferentes a los originales, manteniendo la capacidad de representación de las relaciones empíricas.

Un buen indicador de la utilidad de las transformaciones lo podemos encontrar en el hecho de que prácticamente todos los progra-

mas informáticos de tratamiento de datos (paquetes específicamente estadísticos, hojas de cálculo, gestores de base de datos, etc.) tienen implementadas multitud de funciones y utilidades destinadas a facilitar su realización. Así, por ejemplo, SPSS o SAS tienen alrededor de media docena de instrucciones exclusivamente dedicadas a la transformación de datos, además de múltiples funciones de todo tipo (matemáticas, lógicas, de gestión de valores *missing*, etc.) útiles para las transformaciones de datos (v. gr., Domènech, 1999; Norusis, 2011).

El objetivo de las transformaciones de datos es conseguir alguna ventaja en los análisis, preservando a la vez la información relevante y no dificultando –mejor dicho, en general, facilitando– la interpretación de los resultados. Como veremos, existen numerosas transformaciones, que van desde una operación aritmética (multiplicación, división, etc.) hasta las puntuaciones de razón, la inversa, las escalas de potencia de Tukey (1977), la logarítmica, etc. (v. gr., Freixa, Salafranca, Guàrdia, Ferrer y Turbany, 1992; Peña y Romo, 1997).

Algunas transformaciones permiten, además de simplificar los cálculos, aumentar el nivel de potencia de las pruebas estadísticas, mejorar la simetría de la distribución, comparar valores de distribuciones distintas, etc., lo que facilita la aplicabilidad de las pruebas estadísticas con las que usualmente se analizan los datos (v. gr., Domènech, 2000; Domènech y Sarriá, 1995; Fornieles, Cosculluela y Turbany, 2007; Freixa et al., 1992). Además muchas veces, en caso de que interese, permiten acercar la forma de una relación no lineal a una recta al modificar el escalado de la variable o variables (Moore, 1993; Osorio y Fornieles, 1995; Levin, 1998).

Para que las transformaciones transmitan fielmente la información contenida en los datos originales, Freixa et al. (1992) señalan que es preciso que cumplan la siguiente serie de requisitos:

- 1) *Simplicidad*: propiedad que no se refiere a las operaciones matemáticas implicadas si no que trata sobre el efecto que la transformación ejerce en los datos originales. Las más simples son las lineales, ya que solo afectan al valor numérico. Las transformaciones monótonas no lineales además alteran las distancias relativas, pero conservan el orden de las observaciones originales.
- 2) *Continuidad*: garantiza que solo se producirán los cambios deseables en las distancias relativas entre los puntos de la escala transformada.
- 3) *Monotonidad*: la función debe preservar el orden y, por ende, todos los estadísticos basados en él.
- 4) *Derivabilidad*: asegura la ausencia de brusquedades susceptibles de invalidar la re-expresión.

Siguiendo a Peña y Romo (1997) clasificaremos las transformaciones en lineales y no lineales. Como ya hemos señalado, cuando el objetivo de las transformaciones se limita a facilitar la interpretación de los datos, sin producir ningún cambio esencial en la configuración de la variable, utilizaremos las transformaciones lineales. Estas transformaciones son las más sencillas y las más empleadas en la vida cotidiana (v. gr., la transformación de pesetas a euros).

Hemos visto cómo las transformaciones lineales cambian los valores numéricos de la variable, pero no modifican la forma (asimetría y curtosis) de la distribución. Sin embargo, en muchas ocasiones los datos recogidos en su métrica original presentan valores alejados, asimetría, muestran un patrón de relación no lineal entre variables, etc., que conviene corregir. Las transformaciones no lineales permiten corregir la forma de la distribución, acercando a la normalidad las distribuciones asimétricas y/o con curtosis excesiva. En nuestra exposición nos limitaremos a las transformaciones monótonas, es decir, a aquellas en las que todos los valores de la variable original disminuyen o aumentan su valor aun cuando, obviamente, el montante sea diferente a lo largo del recorrido de la variable.

Centrándonos en primer lugar en el caso univariable, tendrá sentido aplicar una de estas transformaciones cuando queramos convertir en simétrica una distribución que con la métrica original no lo es.

La conveniencia de trabajar con distribuciones simétricas queda justificada por las siguientes razones: a) las distribuciones simétricas permiten describir el centro sin ambigüedades ya que los índices de tendencia central coincidirán, b) además este tipo de distribuciones son más fácilmente interpretables y c) para preservar la robustez de los métodos estadísticos habituales es preciso que los datos sean simétricos (Freixa et al., 1992).

Por último, cabe señalar que en el caso de tratar la cuestión desde el punto de vista de la relación entre variables –por ejemplo, desde la perspectiva de la regresión– las transformaciones permiten rectificar relaciones que no son lineales con los datos originales (v. gr., CEINSA, 1995; Freixa et al., 1992; Moore, 1993; Levin, 1998). En este trabajo también se abordará esta cuestión.

Método

Participantes

La muestra está compuesta por 381 directores técnicos que trabajan en empresas colaboradoras de CEINSA. Para la elaboración de este trabajo se eliminaron previamente los valores extremos de la distribución salarial (por encima del centil 99 y por debajo del centil 1) y los empleados de este puesto que trabajaban en empresas con una facturación mayor de 60.000 millones de pesetas al año (360 millones de euros).

Materiales

La recogida de información para la elaboración del informe se realiza mediante un cuestionario (en papel o en disquete) cumplimentado por las empresas colaboradoras en el que se recoge información sobre las características de la empresa y del empleado necesarias para realizar el modelo de regresión. El tratamiento de los datos se realiza con SPSS y con programas “ad hoc”.

Procedimiento

La recogida de información se realiza enviando por correo convencional o electrónico el cuestionario (ya sea en papel o informatizado) a la empresa durante los meses de marzo y abril.

El plazo de recogida finaliza el 25 de junio y a partir de finales de mayo se lleva a cabo un seguimiento telefónico a las empresas colaboradoras. Las compañías que participan en el estudio reciben información sobre los incrementos salariales y tienen una bonificación del 50% en la compra del informe de remuneraciones. El tratamiento de los datos se lleva a cabo durante los meses de junio y julio, ya que el informe debe ponerse a la venta en el mes de septiembre.

Resultados

A continuación veremos diferentes alternativas de transformación que permiten obtener una mejor simetría de la variable salario. Lo dicho para esta variable es aplicable para la facturación, cuyo comportamiento es todavía más extremo. Estudiaremos en cada caso la asimetría y la curtosis de la variable estudiada.

La gráfica de la figura 2 muestra una distribución salarial clásica (v. gr., CEINSA, 1995, 2012; Moore, 1993; PE, 2006; Peña y Romo, 1997; Watson Wyatt, 2004), con muchos salarios en la parte baja de la distribución y pocos en la parte alta, lo que conlleva que nos encontremos ante una distribución claramente asimétrica (coeficiente de asimetría de Pearson = 1.24, *SE* asimetría = 0.12) y leptocúrtica (coeficiente de curtosis = 1.90, *SE* curtosis = 0.25). Este efecto es todavía más exagerado en la variable ventas (asimetría = 5.99, curtosis = 46.82).

Si deseamos mejorar estos aspectos mediante una transformación la solución ha de pasar por “estirar” los datos hacia la cola de la distribución. Una de las formas de lograrlo es cambiar el escalado de la variable, alargando las distancias entre los puntos. De las posibles alternativas para aplicar en este caso (inversa, logarítmica, raíz cuadrada, etc.) la transformación logarítmica es la que mejor consigna sus objetivos (coeficiente de asimetría = 0.3 y curtosis = 0.2). En la figura 3 se pueden apreciar los resultados obtenidos la transformación logarítmica en base 10, así como también con otras transformaciones.

La primera gráfica (la de mayor tamaño) es la obtenida con la transformación logarítmica en base 10. Las gráficas pequeñas corresponden, por orden, a la transformación realizada al elevar los valores al cuadrado que, como era de esperar, empeora la simetría (coeficiente de asimetría = 2.33). Las dos gráficas siguientes corresponden

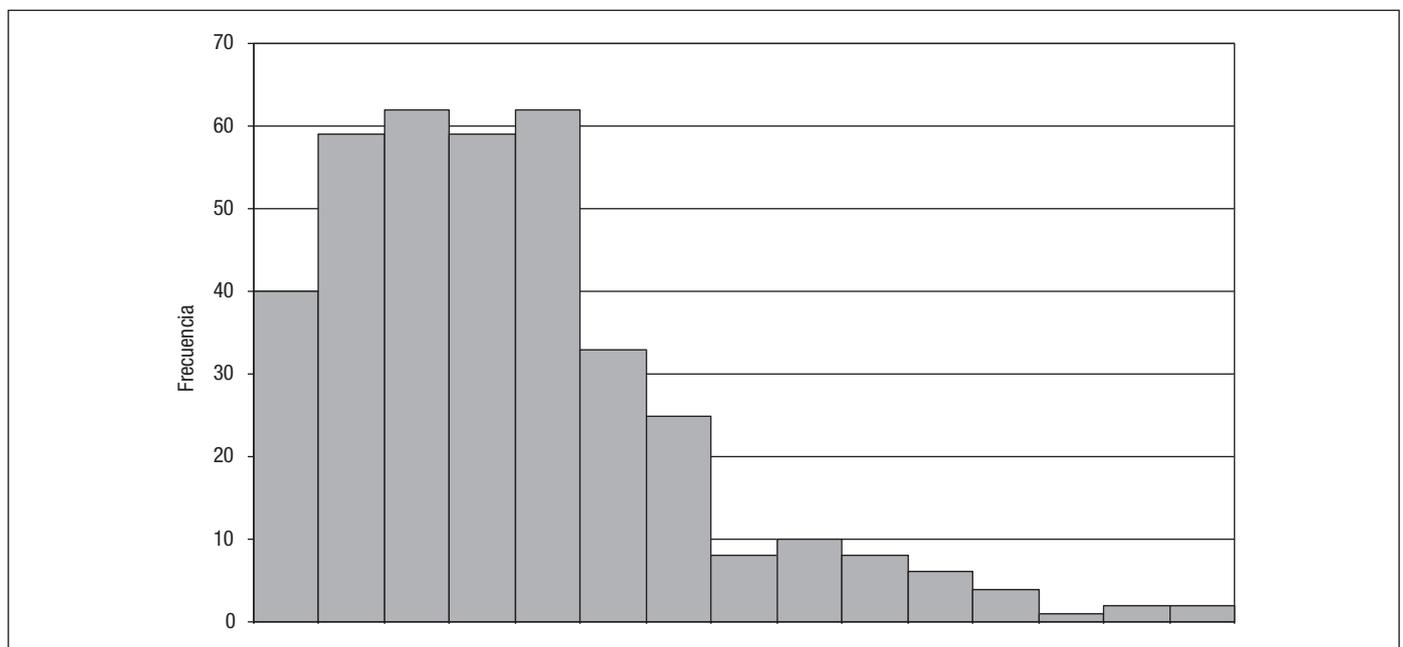


Figura 2. Histograma de los datos salariales originales

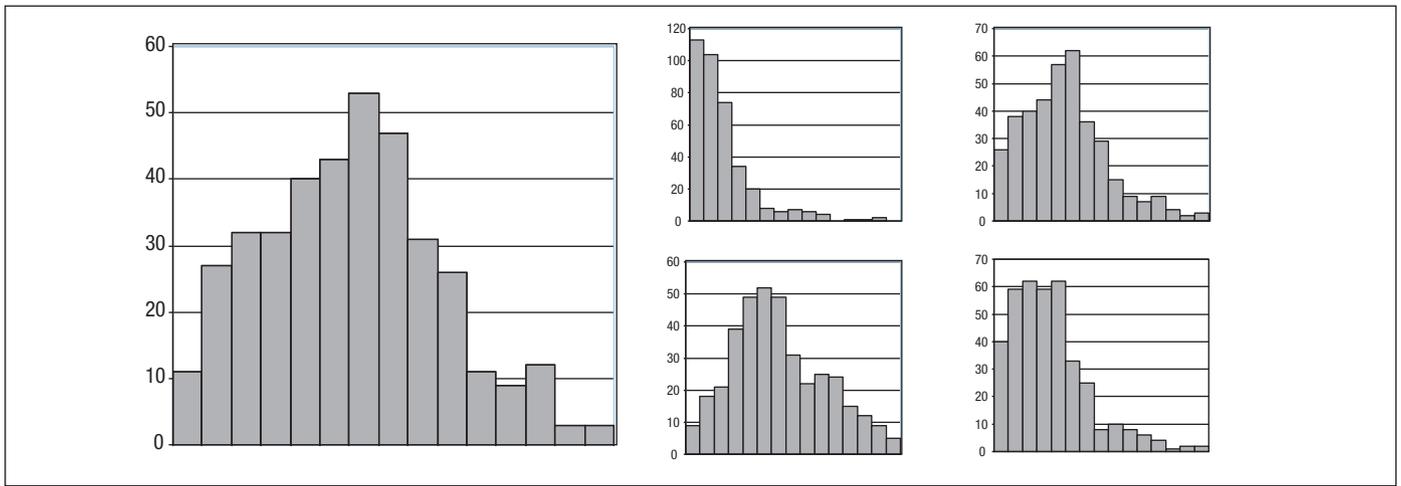


Figura 3. La gráfica de mayor tamaño corresponde a la transformación logarítmica. Las pequeñas corresponden, por orden, a las siguientes transformaciones: elevar los valores al cuadrado, “raíz cuadrada”, “inversa” y “normalizada”.

a las transformaciones “raíz cuadrada” e “inversa”, que mejoran la forma de la distribución, aunque en menor grado que la logarítmica (coeficientes de asimetría = 0.75 y 0.44, respectivamente). Por último, como ejemplo de transformación lineal, hemos incluido la “normalizada”, que obviamente no modifica ni la forma de la distribución ni sus valores de asimetría y curtosis.

Para el caso de la relación entre variables presentamos un ejemplo de la aplicación del modelo de regresión –en este caso simple– para el estudio de la relación entre ventas y salario.

Como ya hemos señalado, una vez que nos situemos en un determinado puesto de trabajo, el tamaño de la organización –operativizado mediante la variable facturación– es la variable que muestra mayor influencia sobre la remuneración.

El volumen de ventas de la empresa es la variable más importante. Para ilustrar su influencia, presentamos en las páginas siguientes un ejemplo de la técnica de regresión, referida en este caso a una sola variable. Se trata, por tanto, de una regresión simple; sin embargo, los principios subyacentes son esencialmente los mismos de la regresión múltiple. En los informes salariales de CEINSA, para las estimaciones salariales utilizamos la técnica de regresión múltiple. Puede verse el modelo completo en CEINSA (2012).

La gráfica de la figura 4 es la clásica “nube de puntos” –diagrama de dispersión–, que representa la relación entre las dos variables. Cada punto simboliza la posición de un individuo. Puede apreciarse cómo la influencia del volumen de ventas de la empresa en la remuneración es evidente. Sin embargo, no se trata de una relación de tipo lineal, lo que aconseja llevar a cabo alguna transformación de las variables que “linealice” la relación.

Las coordenadas de la citada gráfica vienen definidas en sentido horizontal por la cifra de ventas de la empresa en la que trabaja y en el sentido vertical por la remuneración bruta anual que percibe. Como podemos apreciar, se produce una fuerte aglomeración en la zona inferior izquierda. Esta concentración se produce por dos motivos. Tanto la distribución salarial como el tamaño de las empresas tienen –en nuestra muestra y en la realidad– una distribución de tipo piramidal: a medida que aumentamos los valores, disminuye el número (lo que genera la asimetría que hemos visto anteriormente). El formato de presentación utilizado no corrige esta situación, ya que estamos trabajando con escalas lineales, construidas según una progresión aritmética. Como veremos en los cuadros siguientes, la utilización de escalas construidas con otro tipo de progresión produce resultados sensiblemente distintos.

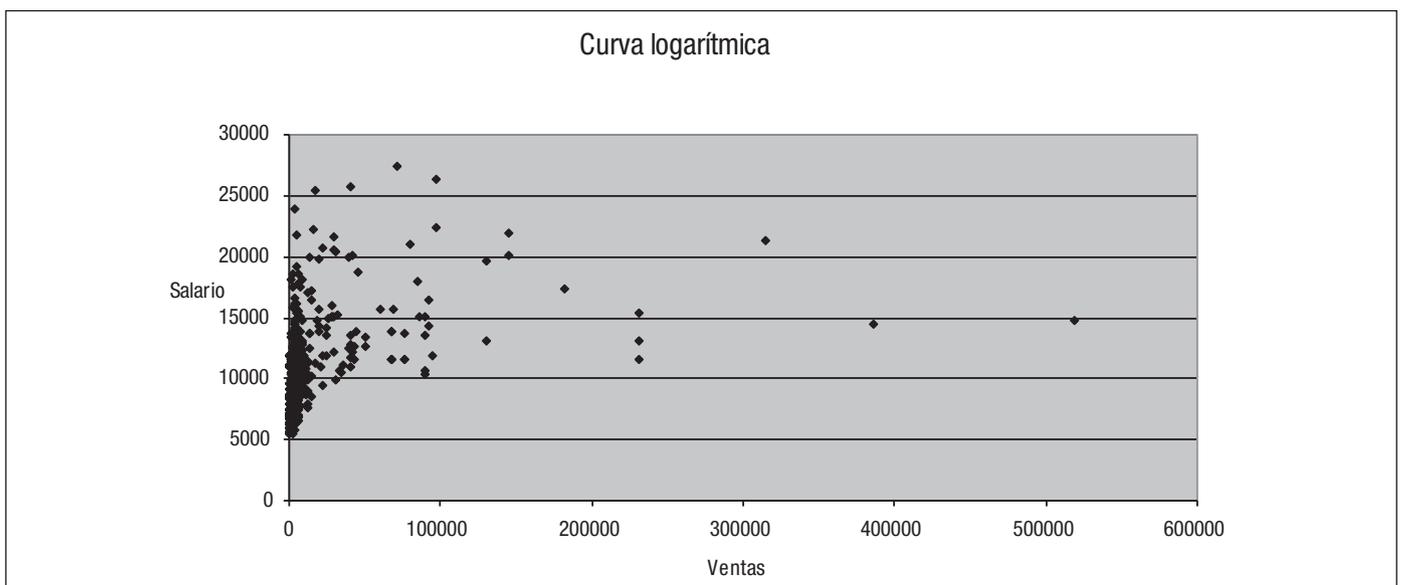


Figura 4. Nube de puntos correspondiente a la relación entre volumen de ventas y salario

En una primera impresión ya se aprecia que los puntos no se hallan distribuidos al azar, siendo apreciable la tendencia a que la remuneración crezca a medida que aumenta el tamaño de la empresa.

En segundo lugar, en la gráfica de la figura 5 hemos trazado una "línea de tendencia" en la que se puede apreciar claramente una curvatura progresivamente desacelerada. Esta forma curvilínea que aparece en la figura –con una pendiente pronunciada al principio, que después se va suavizando progresivamente– es característica de la relación entre el tamaño de la empresa y la remuneración de los cargos ejecutivos.

Queda claro que si intentáramos representar la relación entre ambas variables por medio de una recta, obtendríamos un mal ajuste.

Por ejemplo, si construyéramos la recta sobre el tramo inicial que agrupa el grueso de las observaciones, en la derecha "nos saldríamos del cuadro", es decir, su prolongación produciría remuneraciones disparatadas para los individuos que trabajan en empresas grandes.

Aunque a primera vista puede dar una impresión distinta, la distribución que aparece en la gráfica de la figura 6 es exactamente la misma de las dos gráficas anteriores. La diferencia estriba en que en este caso utilizamos escalas logarítmicas para representar ambas variables.

Aunque la línea de tendencia que vimos en la gráfica anterior se puede utilizar para estimar la remuneración correspondiente a un determinado tamaño de empresa, se utiliza una recta de regresión

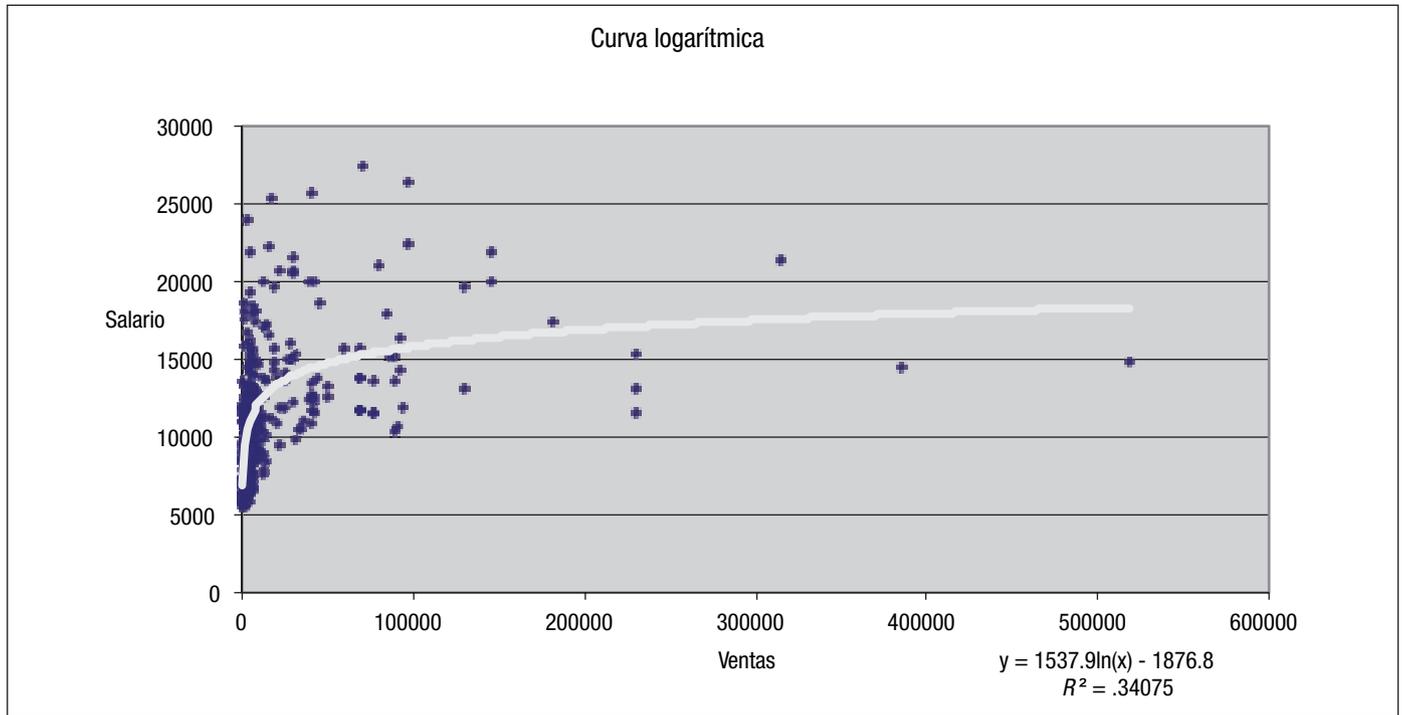


Figura 5. Línea de tendencia de relación entre volumen de ventas y salario

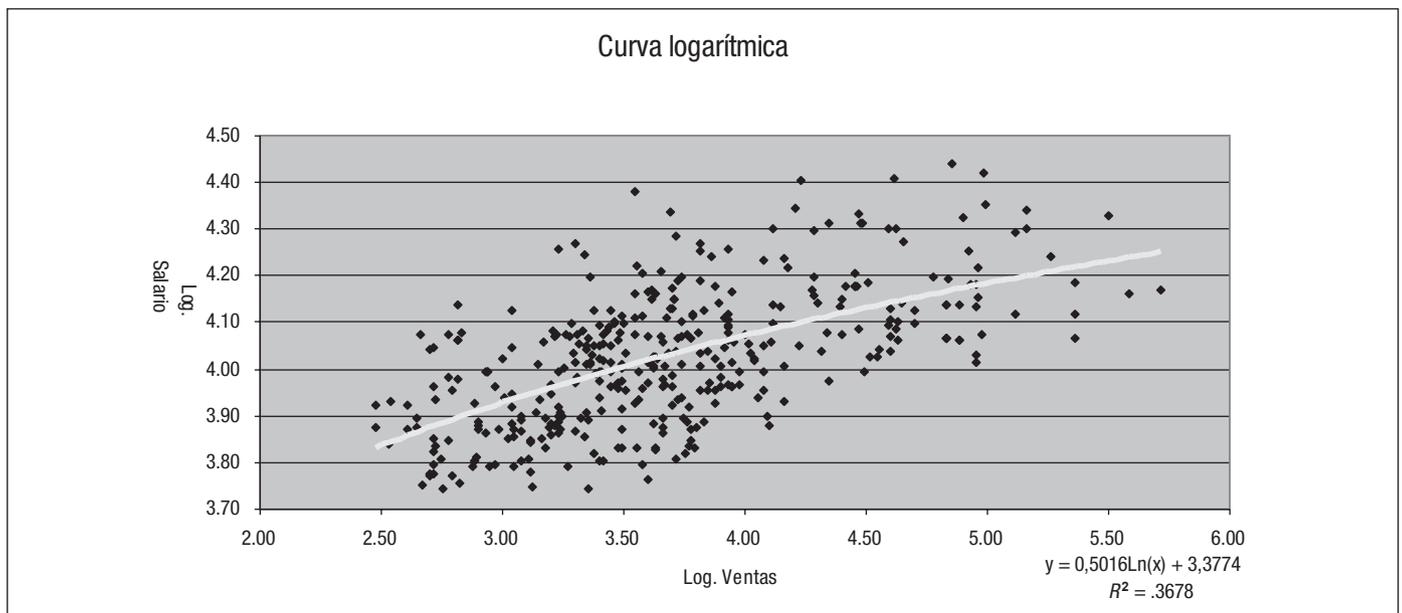


Figura 6. Línea de tendencia de la relación entre los logaritmos del volumen de ventas y del salario

como la que aparece representada en la gráfica de la figura 7, que puede ser expresada mediante una expresión matemática sencilla $Y = K + b \cdot X$.

La pendiente de la recta pone de manifiesto una estrecha relación entre el tamaño de la empresa y la remuneración, característica de los puestos directivos. Si trazáramos las rectas de regresión para los diferentes puestos, observaríamos que la influencia del tamaño se va atenuando a medida que descendemos a través de la escala jerárquica.

Discusión

Hemos visto como, tras probar diferentes alternativas, la transformación logarítmica es la opción que mejor corrige la asimetría y la curtosis de la variable estudiada, consiguiendo un acercamiento más que aceptable a una distribución de tipo normal. Además, por lo que respecta a la relación entre variables, permite convertir en lineal una relación que con los datos originales no lo es.

Esto es consecuencia directa del hecho de que en una escala lineal la distancia entre dos cantidades es proporcional al valor directo de éstas. Por ejemplo, refiriéndonos al volumen de ventas, la distancia entre 5.000 y 10.000 sería la misma que entre 35.000 y 40.000. Por el contrario, en una escala logarítmica, la distancia entre dos cantidades es proporcional al logaritmo de éstas. Por ejemplo, entre 1.000 ($\log = 3$) y 10.000 ($\log = 4$) existe la misma distancia que entre 10.000 ($\log = 4$) y 100.000 ($\log = 5$). La gráfica adjunta no llega hasta 100.000, pero se puede observar esta equivalencia, por ejemplo, entre 200 y 300, 2.000 y 3.000, 20.000 y 30.000, etc.

La primera consecuencia de la transformación logarítmica es que la línea de tendencia que antes se acercaba a una curva potencial ahora se aproxima razonablemente a la forma de una línea recta. Otra consecuencia secundaria es que la aglomeración de puntos en el extremo inferior izquierdo de la gráfica se distribuye ahora de manera más uniforme, lo que nos permite recuperar la representación gráfica de la nube de puntos, con una discriminación aceptable. La pendiente de la recta pone de manifiesto una estrecha relación entre tamaño de empresa y remuneración, característica de puestos directivos. Si trazáramos las rectas de regresión para los diferentes pue-

tos, observaríamos que la influencia del tamaño se va atenuando a medida que descendemos a través de la escala jerárquica.

En definitiva, la transformación logarítmica es la que mejor cumple con los propósitos de este trabajo, tanto en el caso univariable como en relación entre variables.

Extended Summary

The present work shows some aspects related with the treatment of data collected from salary surveys. The use of transformations has allowed us to improve the possibility of analysis and interpretability of the data. Along the article we will study the relationship between salary and size of the company (turnover). Clearly, turnover is the more effective salary predictor. However, the relationship between turnover and salary is not linear. If we want to use linear regression, variables must be transformed.

A good indicator of the utility of the transformations can be found in the fact that all software for data analysis has implemented several functions to facilitate its performance. In this case, the aim of the transformations is to achieve some advantage in the analyses, facilitating the interpretation of the results. More specifically, the aim is to move the form of a no linear relation closer to a straight.

The transformations could be classified in linear and no linear. When the aim of the transformations limits the interpretation of the data without producing any essential change in the configuration of the variable, linear transformations will be used. These transformations are the simplest and the most commonly applied in daily life (e.g., to transform US Dollars to Euros). Linear transformations change values, but do not modify the form (asymmetry and kurtosis) of the distribution. However, frequently the data collected in his original metrics present asymmetry and show a pattern of no linear relation between variables. The non linear transformations allow correcting the form of the distribution, drawing the asymmetric distribution near normality.

First, we will begin with a single-variable case; it will make sense to apply one of these transformations when we want to turn symmetrical a distribution that is not symmetrical in the original

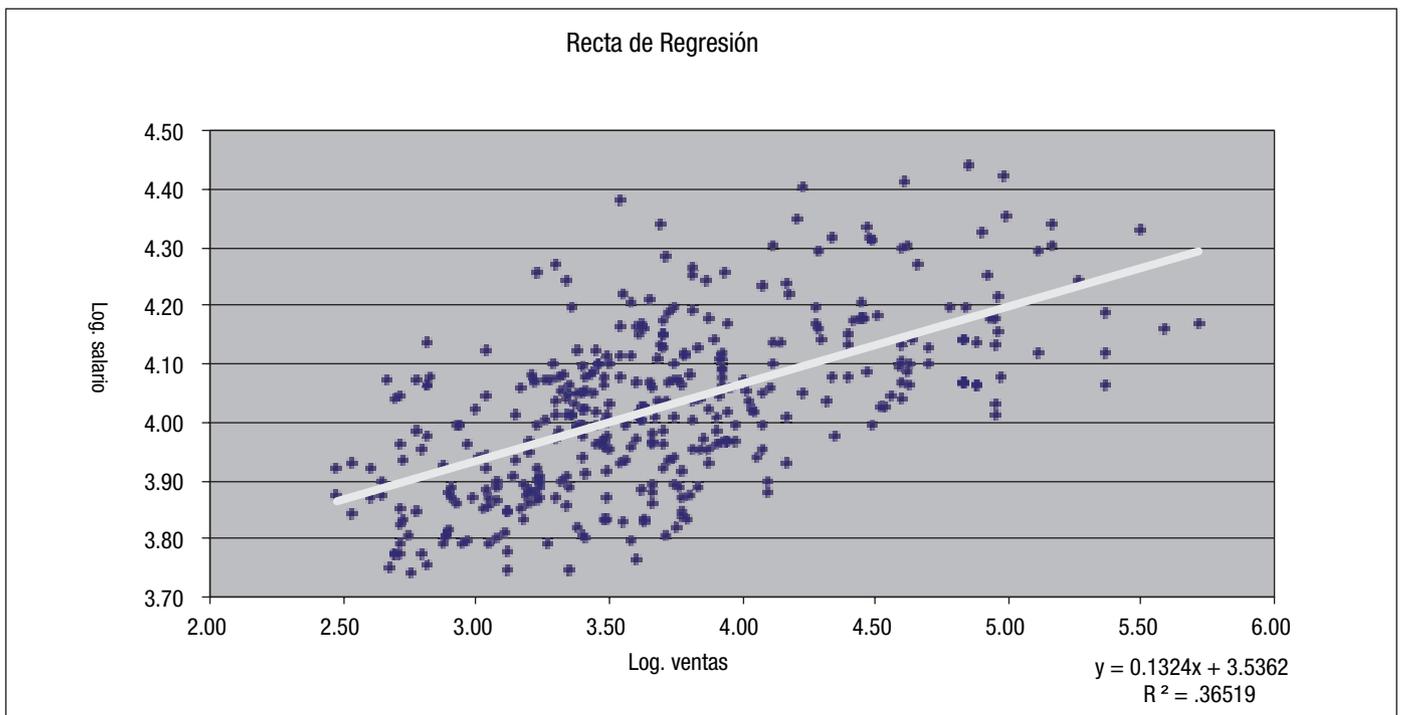


Figura 7. Recta de regresión entre los logaritmos del volumen de ventas y del salario

metrics. The advantage of working with symmetrical distributions is justified by the following reasons: a) the symmetrical distributions allow us to describe the centre without ambiguities; b) moreover, this type of distributions are more easily interpretable; and c) in order to preserve the robustness of the usual statistical methods the data must be symmetrical. Finally, we will address the question of the relationship between variables from the perspective of linear regression; the transformations try to straighten relations that are not linear in the original data. In this work we will also tackle this question.

Method

Sample and procedure

The sample is composed of 381 Technical Directors working in companies partners of CEINSA (company of Human Resources specialised in salary reports and studies) who filled in some questionnaires. For the preparation of this work the extreme values of the salary distribution (above centile 99 and below centile 1) and the employees working in companies with a turnover exceeding 360 million euros were excluded. The questionnaire contains the necessary information about the company and the employees to perform the model of regression (Company: turnover, geographic location, number of employees, sector of activity, etc.; Employee: total remuneration, variable remuneration, age, seniority, level of education, number of subordinates, etc.).

Results and Discussion

We will try different alternatives of transformation in order to obtain a better symmetry of variable wage. The same treatment is applied for the turnover, a variable that behaves even more extremely. In each case, we will study the asymmetry and the kurtosis of the variable studied.

Figure 2 shows a classical salary distribution, with most wages concentrated at the bottom and a few at the top. We are consequently confronted clearly with an asymmetric (Pearson's coefficient of asymmetry = 1.24, *SE* asymmetry = 0.12) and leptokurtic distribution (kurtosis = 1.90, *SE* kurtosis = 0.25). This effect is even more exaggerated in the variable sales. (asymmetry = 5.99, kurtosis = 46.82).

If we wish to improve these effects by means of a transformation, one of the ways to do it is to change the scale of the variable, lengthening the distances between the points. Possible alternatives to apply in this case are reverse, logarithmic, square root transformations. The logarithmic transformation is the best to achieve the goal (asymmetry = 0.3 and kurtosis = 0.2).

The first graphic in figure 3 (the largest) is the one obtained with the logarithmic transformation to base 10. The other graphics correspond to the transformation performed when increasing the values to the square, square root, and inverse transformations, that improve the form of the distribution, although to a lower degree than the logarithmic one (asymmetry = 0.75 and 0.44, respectively). Finally, as an example of linear transformation, we have included the normal one. Obviously, it does not modify the shape of the distribution (same values of asymmetry and kurtosis).

In the case of the relationship between the turnover and wage variables, an application of a simple regression model is presented. As we have already highlighted, the size of the company (turnover) is the variable that shows a greater influence on the remuneration (in CEINSA's salary reports, for estimated market salaries we use multiple regression models).

Figure 4 shows the scatter diagram, that graphically represents the relation between both variables. We can observe how the influence of turnover over remuneration is evident. However, it is

not a linear relationship, which would recommend carrying out some transformation.

The coordinates of the afore-mentioned graphic are defined horizontally by the amount of sales of the company in which the employee works and vertically by the yearly gross income the employee gets. As we can see, it produces a strong agglomeration in the lower left zone. The format of the presentation used does not correct this situation, since we are working with linear scales, built according to an arithmetical progression.

As we will see in the following graphics, the use of scales built with another type of progression produces noticeably different results. It can firstly be observed that the points are not found randomly distributed, showing a tendency of the remuneration to grow at the same pace as the turnover does.

In Figure 5 we have plotted a trend line, in which an increasingly decelerated curvature can be noted. This curvilinear shape that appears in the figure with a pronounced slope at the beginning, progressively softening, is characteristic of the relation between the size of the company and the managers' income.

It remains clear that if we tried to represent the relationship between both variables by means of a straight line, we would obtain a poor adjust. For example, if we built the straight line on the initial stretch that groups the thickness of the observations, it would go out of the picture on the right side, that is to say, its prolongation would produce absurd remuneration for individuals who work in big companies.

Although at first sight it would not be noticed, the distribution that appears in Figure 6 is exactly the same as the two previous graphics. The difference is that in this case we use logarithmic scales to represent both variables.

Although the trend line observed in the previous graphic can be used to estimate the corresponding remuneration of a certain size of company, a straight line on the regression, like the one represented in figure 7, that can be expressed by means of a simple mathematical expression $Y = K + b * X$, is normally used instead.

The slope ("b" of the equation) of the straight line evidences a narrow relation between size of the company and remuneration, characteristic of the managerial positions. If we plotted the straight line on the regression for the different positions, we would observe that the influence of the size diminishes as we go down through the hierarchical scale.

After testing different alternatives, the logarithmic transformation is the best option to correct the asymmetry and the kurtosis, achieving a more than acceptable approach to a normal distribution. Besides, considering the relationship between variables, it allows turning a relation linear, which is not so in the original data.

This is a direct consequence of the fact that in a linear scale the distance between two quantities is proportional to the direct value of these. For example, regarding the volume of sales, the distance between 5,000 and 10,000 would be the same as between 35,000 and 40,000. On the contrary, in a logarithmic scale, the distance between two quantities is proportional to the logarithm of these. For example, between 1,000 ($\log = 3$) and 10,000 ($\log = 4$) the same distance exists as between 10,000 ($\log = 4$) and 100,000 ($\log = 5$). The graphic shown does not reach 100,000, but this equivalence can be observed, for example, between 200 and 300, 2,000 and 3,000, 20,000, and 30,000.

The first consequence of the logarithmic transformation is that the trend line that almost got into a potential curve earlier, gets now reasonably closer to the form of a straight line. Another consequence is that the agglomeration of points on the lower left extreme of the graphic is distributed in a more uniform way, which allows us to recover the graphic representation of the cloud of points, with an acceptable discrimination. The slope of the straight line evidences a close relationship between size of company and remuneration.

In conclusion, the logarithmic transformation is the one which best fulfils the purposes of this work, as much in the case of single-variable as in the relationship between variables.

Conflicto de intereses

El autor de este artículo declara que no tiene ningún conflicto de intereses.

Referencias

- CEINSA (1995). *Políticas retributivas*. Barcelona: Ceinsa.
- CEINSA (2012). *Informe de Remuneraciones*. Barcelona: Ceinsa.
- Domènech, J. M. (1999). *Introducción al paquete estadístico SPSS en Ciencias de la Salud*. Bellaterra: Universitat Autònoma de Barcelona.
- Domènech, J. M. (2000). *Diagnósticos de una modelo de regresión múltiple. Análisis multivariante: Modelos de Regresión*. Barcelona: Signo.
- Domènech, J. M. y Sarriá, A. (1995). *Métodos de estimación. Análisis multivariante: Modelos de Regresión*. Barcelona: Signo.
- Fornieles, A., Cosculluela, A y Tubany, J. (2007). *Análisis de Datos en Psicología*. Barcelona: Editorial UOC.
- Freixa, M., Salafranca, Ll. Guàrdia, J., Ferrer, R. y Turbany, J. (1992). *Análisis Exploratorio de Datos: Nuevas técnicas estadísticas*. Barcelona: PPU.
- Levin, R. I. (1998). *Estadística para administradores (7ª ed.)*. New Jersey, NJ: Prentice.
- Moore, D. (1993). *Decisions Through Data*. Colección de vídeos. Barcelona: Fundació per la UOC.
- Norusis. (2011). *SPSS Base 19.0*. Chicago: Autor.
- PE (2006). *Remunerations survey*. London: PE.
- Peña, D. y Romo, J. (1997). *Introducción a la Estadística para Ciencias Sociales*. Madrid: McGraw Hill.
- Salvador, F. (1996). *Quantificació de les observacions, escales de mesura*. Barcelona: Edi UOC.
- Tukey, J. W. (1977). *Exploratory Data Análisis*. Reading, MA: Addison-Wesley.
- Watson Wyatt (2004). *Informe de Remuneraciones*. Barcelona: Watson Wyatt.