

REGRESIÓN DE LAS VENTAS SOBRE LOS GASTOS DE PERSONAL EN EMPRESAS DE LA PROVINCIA DE LA CORUÑA

Sánchez Sellero, César¹ y Sánchez Sellero, M. Carmen²

¹Departamento de Estadística e Investigación Operativa
Universidade de Santiago de Compostela

²Departamento de Economía Aplicada II
Universidade da Coruña

RESUMEN

En este trabajo se estudia la relación entre el importe neto de las ventas efectuadas por las empresas y sus gastos de personal. Para ello, se ajusta un modelo de regresión lineal sobre los logaritmos de ambas variables. Diversos métodos de contraste muestran diferencias significativas entre grupos de actividades económicas. Por último, se contrasta la buena especificación del modelo.

1. INTRODUCCIÓN

El propósito de este trabajo es estudiar la relación entre el importe neto de las ventas efectuadas por las empresas y sus gastos de personal. Para ello, construiremos un modelo de regresión que además tenga en cuenta la actividad económica a la que se dedica la empresa.

Ajustaremos el modelo en base a la información proveniente de la base de datos "SABI. Sistema de Análisis de Balances Ibéricos", actualizada a junio del 2003. De esta base de datos hemos tomado las empresas localizadas en la provincia de La Coruña, y en concreto utilizamos las variables: Código CNAE 93 (clasificación nacional de actividades económicas), importe neto de cifra de ventas en miles de euros del año 2001 y los gastos de personal en miles de euros del mismo año. Excluimos las empresas con algún dato ausente de los anteriores, y también aquellas cuyos gastos de personal son superiores a su cifra de ventas, pues presentan un comportamiento distinto al de las demás empresas. Finalmente, el estudio se efectuó con los datos de 6875 empresas.

Se consideró el código CNAE como instrumento para identificar la actividad económica, y se agregaron los códigos CNAE en grupos homogéneos, como se verá más adelante. En cualquier caso, parece natural pensar que la relación entre las ventas y los gastos de personal dependerá de la actividad económica de la empresa. Este hecho se constatará en el presente trabajo al ajustar el modelo de regresión.

Los últimos datos disponibles relativos a las ventas y los gastos de personal se refieren al año 2001, ya que las empresas han de presentar cada año las cuentas del año anterior (en el 2003 las cuentas del 2002), por lo que en esta base de datos actualizada a junio del 2003 sólo se encuentran los datos del 2001.

En Amat (2002) se podrá encontrar un tratamiento detallado del significado, utilidad y limitaciones de los análisis de estados financieros de las empresas y de los estudios sectoriales.

2. CONSTRUCCIÓN DEL MODELO

Se formaron grupos de actividades económicas agregando códigos CNAE similares. Los grupos formados figuran en la tabla siguiente:

Códigos CNAE	Actividad económica
01, 02	Agricultura, ganadería, caza y silvicultura
05	Pesca y acuicultura
10, 11, 12	Extracción de productos energéticos
13, 14	Extracción de otros minerales, excepto productos energéticos
15, 16	Industria de la alimentación, bebidas y tabaco
17, 18	Industria textil, de la confección y de la peletería
19	Industria del cuero y del calzado
20	Industria de la madera y del corcho
21, 22	Industria del papel; edición, artes gráficas y reproducción de soportes grabados
24	Industria química
25	Fabricación de productos de caucho y materias plásticas
26	Fabricación de otros productos minerales no metálicos
27, 28	Metalurgia y fabricación de productos metálicos
29	Industria de la construcción de maquinaria y equipo mecánico
30, 31, 32, 33	Fabricación de material y equipo eléctrico, electrónico y óptico
34, 35	Fabricación de material de transporte
36, 37	Fabricación de muebles y otras industrias manufactureras. Reciclaje
40, 41	Producción y distribución de energía eléctrica, gas y agua
45	Construcción
50, 51, 52	Comercio. Reparación de vehículos de motor, y de efectos personales y de uso doméstico
55	Hostelería
60, 61, 62, 63, 64	Transporte, almacenamiento y comunicaciones
65, 66, 67	Intermediación financiera
70, 71, 72, 73, 74	Actividades inmobiliarias y de alquiler; servicios empresariales
80	Educación
85	Actividades sanitarias y veterinarias, servicios sociales
90, 91, 92, 93	Actividades de saneamiento público, asociativas, recreativas, culturales, deportivas y de servicios personales

Dado que el gasto de personal y el importe de las ventas son variables positivas y con un marcado comportamiento asimétrico, hemos efectuado una transformación logarítmica a cada una de ellas, para poder recurrir después a procedimientos estadísticos relativamente comunes, pero que requieren variables simétricas y en muchos casos adaptables al modelo de distribución normal.

Denotemos P_{ij} , V_{ij} a los gastos de personal y el importe de las ventas, respectivamente, de la empresa j en el grupo de actividades económicas i , donde $j \in \{1, \dots, n_i\}$ $i \in \{1, \dots, I\}$.

Entonces planteamos el modelo de regresión lineal

$$\log V_{ij} = \alpha_i + \beta_i \log P_{ij} + \varepsilon_{ij} \quad j \in \{1, \dots, n_i\} \quad i \in \{1, \dots, I\}$$

donde α_i, β_i son los parámetros del modelo, que se han de estimar, y ε_{ij} son variables con distribución normal de media cero y varianza σ_i^2 , e independientes entre sí. Aplicando una transformación exponencial podemos ver este modelo en la forma,

$$V_{ij} = e^{\alpha_i} \cdot P_{ij}^{\beta_i} \cdot e^{\varepsilon_{ij}} \quad j \in \{1, \dots, n_i\} \quad i \in \{1, \dots, I\}$$

de manera que se trata de un modelo potencial, en el cual el error se incorpora de forma multiplicativa y tiene distribución lognormal. Este modelo es apropiado para este tipo de situaciones, por otro lado muy frecuentes en la economía. Nótese que si $\beta_i = 1$, estamos ante un

modelo lineal, y en tal caso e^{α_i} constituye la pendiente de la recta, la cual se puede interpretar como un ratio entre el importe de las ventas y los gastos de personal. Explicaciones más detalladas sobre este tipo de ratios se pueden encontrar en Banegas Ochovo y otros (1998).

A continuación, procedemos a estimar los parámetros. Para ello, introducimos la notación

$$Y_{ij} = \log V_{ij} \quad X_{ij} = \log P_{ij}$$

y entonces los estimadores de los parámetros del modelo se pueden expresar

$$\hat{\beta}_i = \frac{S_{XYi}}{S_{Xi}^2} \quad \hat{\alpha}_i = \bar{Y}_i - \hat{\beta}_i \bar{X}_i \quad SRC_i = \sum_{j=1}^{n_i} (Y_{ij} - \hat{\alpha}_i - \hat{\beta}_i X_{ij})^2 \quad \hat{\sigma}_i = \sqrt{SRC_i / (n_i - 2)}$$

$$\text{donde } \bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij} \quad \bar{Y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij} \quad S_{Xi}^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2 \quad S_{XYi} = \frac{1}{n_i} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)(Y_{ij} - \bar{Y}_i)$$

Nótese que estamos efectuando la estimación de un modelo de regresión lineal en cada grupo de manera independiente. Los resultados están contenidos en la tabla siguiente:

Códigos CNAE	n_i	$e^{\hat{\alpha}_i}$	$\hat{\beta}_i$	$\hat{\sigma}_i$
01, 02	59	19'536	0'721	0'915
05	80	6'632	0'846	0'646
10, 11, 12	4	3'529	1'043	1'084
13, 14	24	2'581	1'049	0'472
15, 16	170	3'137	1'108	0'813
17, 18	134	2'762	1'036	0'928
19	5	269'951	0'372	1'348
20	135	3'685	1'059	0'691
21, 22	135	7'756	0'847	0'645
24	21	3'029	1'124	0'867
25	31	6'732	0'993	0'654
26	102	1'850	1'171	0'687
27, 28	279	2'449	1'087	0'582
29	48	4'105	1'003	0'649
30, 31, 32, 33	29	4'328	0'991	0'747
34, 35	46	2'782	1'012	0'585
36, 37	101	3'044	1'017	0'653
40, 41	34	94'974	0'567	1'323
45	1372	4'064	0'945	0'679
50, 51, 52	2363	9'592	0'967	0'866
55	202	5'270	0'892	0'396
60, 61, 62, 63, 64	363	9'695	0'849	0'810
65, 66, 67	41	3'392	0'998	0'637
70, 71, 72, 73, 74	857	16'543	0'710	1'107
80	48	9'139	0'725	0'665
85	50	5'659	0'889	0'532
90, 91, 92, 93	142	6'504	0'902	0'838

Observamos que en muchos casos el coeficiente $\hat{\beta}_i$ toma un valor próximo a uno, de modo que el modelo se asemeja a una recta de las ventas sobre los gastos de personal, cuya pendiente $e^{\hat{\alpha}_i}$ constituye un ratio entre ambas variables. Aunque a simple vista, parece haber diferencias notables en los resultados para los distintos grupos, vamos a efectuar contrastes sobre la igualdad entre grupos. La primera hipótesis de este tipo sería que la varianza del error es la misma en todos los grupos:

$$H_0 : \sigma_1^2 = \dots = \sigma_I^2 (= \sigma^2)$$

El test de razón de verosimilitudes para contrastar esta hipótesis nula conduce al estadístico

$$\sum_{i=1}^I n_i \log \frac{\text{SRC}/n}{\text{SRC}_i/n_i}$$

siendo $\text{SRC} = \sum_{i=1}^I \sum_{j=1}^{n_i} (Y_{ij} - \hat{\alpha}_i - \hat{\beta}_i X_{ij})^2$ la suma residual de cuadrados total y $n = \sum_{i=1}^I n_i$ el tamaño muestral total. Nótese que SRC/n es el estimador de máxima verosimilitud de la varianza común σ^2 , mientras que SRC_i/n_i es el estimador de máxima verosimilitud de la varianza del grupo i-ésimo σ_i^2 . Aproximamos la distribución del estadístico de razón de verosimilitudes por una ji-cuadrado con I-1=26 grados de libertad. El valor que tomó el estadístico fue 612'93, que da lugar a un nivel crítico extremadamente próximo a cero.

Para contrastar la hipótesis nula

$$H_0 : \alpha_1^2 = \dots = \alpha_I^2 \quad \text{y} \quad \beta_1^2 = \dots = \beta_I^2$$

esto es, la igualdad de las rectas de regresión (de sus parámetros) entre los grupos, empleamos el test F, que también resultó muy significativo.

En resumen, todos los intentos de simplificar el modelo resultaron infructuosos, y en consecuencia debemos concluir que los grupos de actividades económicas que hemos considerado son distintos en lo que respecta a la relación entre el importe de las ventas y los gastos de personal. Claro está que todo lo anterior es correcto suponiendo cierto el modelo considerado. Por tanto, el siguiente paso deberá ser comprobar que el modelo es plausible, tarea a la cual se dedica la sección siguiente.

3. CONTRASTE DE ESPECIFICACIÓN DEL MODELO

Se trata de contrastar la hipótesis nula

$$H_0 : Y_{ij} = \alpha_i + \beta_i X_{ij} + \varepsilon_{ij} \quad \text{siendo} \quad \varepsilon_{ij} \in N(0, \sigma_{ij}^2) \quad \text{independientes,}$$

frente a la alternativa de que la función de regresión no se adapte a este modelo. Usaremos un test basado en la función de regresión integrada, que desemboca en un proceso de contraste de la forma

$$R_i(x) = \frac{1}{\sqrt{n_i}} \sum_{j=1}^{n_i} (Y_{ij} - \hat{\alpha}_i - \hat{\beta}_i X_{ij}) I\{X_{ij} \leq x\} \quad \text{con} \quad x \in \mathbb{R}$$

donde el símbolo $I\{\cdot\}$ representa la función indicadora del suceso contenido entre las llaves. Para cada grupo de actividades económicas, representado por el subíndice i, tenemos un proceso R_i , que permite contrastar si en ese grupo se respeta el modelo. Como estadísticos de contraste

podríamos considerar el supremo, $\sup_{x \in \mathbb{R}} R_i(x)$, o un estadístico del tipo Cramer von Mises,

$$\frac{1}{n_i} \sum_{j=1}^{n_i} [R_i(X_{ij})]^2.$$

Para aproximar la distribución del proceso de contraste y de ahí obtener el nivel crítico, aplicamos un procedimiento bootstrap de generación de muestras artificiales, de la forma

$$Y_{ij}^* = \hat{\alpha}_i + \hat{\beta}_i X_{ij} + \varepsilon_{ij}^* \quad \text{siendo } \varepsilon_{ij}^* \in N(0, \hat{\sigma}_{ij}^2) \text{ independientes,}$$

que es muy similar al propuesto en Stute, González Manteiga y Presedo Quindimil (1998), con la salvedad de que los errores bootstrap siguen el modelo normal planteado en la hipótesis nula. A partir de las muestras bootstrap y con las mismas operaciones realizadas en la muestra original, se pueden obtener versiones bootstrap del proceso de contraste y de los estadísticos de contraste. Un recuento de la réplicas bootstrap que toman valores superiores a la versión original del estadístico de contraste conduce al nivel crítico. Los resultados obtenidos tras haber extraído cien mil réplicas bootstrap constan en la tabla siguiente:

Códigos CNAE	Nivel crítico	Nivel crítico
	Estadístico del supremo	Estadístico de Cramer von Mises
01, 02	0'502	0'565
05	0'025	0'045
10, 11, 12	0'512	0'562
13, 14	0'917	0'969
15, 16	0'595	0'515
17, 18	0'00005	0'00045
19	0'430	0'459
20	0'111	0'153
21, 22	0'375	0'198
24	0'753	0'800
25	0'928	0'900
26	0'641	0'687
27, 28	0'855	0'690
29	0'640	0'689
30, 31, 32, 33	0'819	0'743
34, 35	0'859	0'917
36, 37	0'354	0'224
40, 41	0'874	0'889
45	0	0
50, 51, 52	0'021	0'00998
55	0'850	0'898
60, 61, 62, 63, 64	0'904	0'935
65, 66, 67	0'855	0'936
70, 71, 72, 73, 74	0'00871	0'00919
80	0'439	0'205
85	0'679	0'807
90, 91, 92, 93	0'909	0'947
Global	0'00472	0'02296

En la última fila, titulada “Global”, hemos añadido el nivel crítico correspondiente al contraste de que el modelo se cumple en todos los grupos de actividades económicas, contraste que basamos en el máximo de los estadísticos obtenidos para cada grupo. Los niveles críticos permiten aceptar el modelo para casi todos los grupos. Por el contrario, resultan muy significativos los resultados para

la industria textil, de la confección y de la peletería (códigos CNAE 17 y 18) y para la construcción (código 45). También resultan algo significativos, aunque en un orden de magnitud muy distinto, los resultados para las actividades inmobiliarias y de alquiler y los servicios empresariales (códigos 70, 71, 72, 73 y 74), para el comercio y las actividades de reparación (50, 51 y 52) y para la pesca y acuicultura (código 05). En estos casos, habría que indagar si subyace un modelo distinto, paramétrico o no, o si es necesaria una desagregación más detallada del grupo en otros que sí satisfagan el modelo planteado en este trabajo.

Los niveles críticos obtenidos para el contraste global son pequeños, aunque no demasiado, lo cual es coherente con los resultados obtenidos para cada grupo individualmente, pues parece que la mayoría de los grupos sí cumplen el modelo, mientras que unos pocos presentan indicios de no cumplirlo.

4. CONCLUSIONES

Como consecuencia de los análisis realizados, concluimos que un modelo de regresión lineal del logaritmo de las ventas sobre el logaritmo de los gastos de personal se ajusta bastante bien a los datos de las empresas de la provincia de La Coruña, cuando éstas se agrupan por similitud en su actividad económica. Además, este modelo es recomendable en virtud de las propiedades de las variables consideradas, y goza de interpretaciones naturales en el ámbito empresarial.

Bajo el modelo considerado, se constatan diferencias significativas entre grupos de actividades económicas, tanto en los coeficientes de la función de regresión, como en la varianza del error.

5. REFERENCIAS

Amat, O. (2002). *Análisis de estados financieros. Fundamentos y aplicaciones*. Ediciones Gestión 2000, S.A.

Banegas Ochovo, R., Sánchez Mayoral García-Calvo, F. y Nevado Peña, D. (1998). *Análisis por ratios de los estados contables financieros (Análisis externo)*. Editorial Civitas, S.A.

Stute, W., González Manteiga, W. y Presedo Quindimil, M. (1998): "Bootstrap approximations in model checks for regression". *Journal of the American Statistical Association* 93, 141-149.