

MODELO EXPLICATIVO DE LAS VENTAS ANUALES EN EL SECTOR PEQUEÑO COMERCIO DE CÓRDOBA. ANÁLISIS DE LA NO RESPUESTA EN LAS ENCUESTAS DE ESTUDIO DEL SECTOR.

Autores:

Prof. Dra. Mercedes Torres Jiménez.
Prof. Juan Manuel Arjona Fuentes.

Institución/Universidad:

Facultad de Ciencias Empresariales de Córdoba (ETEA). Universidad de Córdoba.

Área Temática:

Economía Industrial y de Servicios.

Dirección de Contacto:

ETEA. C/ Escritor Castilla Aguayo, 4. 14004 Córdoba **Teléfono:** 957222100 **Fax:** 957222101

Correo electrónico: mtorres@etea.com jmarjona@etea.com

Resumen:

En el año 2001 el grupo de investigación sobre pequeño comercio constituido en la Facultad de Ciencias Empresariales (ETEA) de la Universidad de Córdoba, abordaba su segundo Estudio del Comercio en la Provincia de Córdoba.

En dicho estudio, como en otros muchos, se puede observar un alto grado de no respuesta a las preguntas directas sobre magnitudes económicas del negocio, que obliga a los investigadores a extraer conclusiones con un número menor de datos, lo que puede poner en entredicho la fiabilidad de la inferencia, o bien deducirlo a través de respuestas con horquilla de valores, con la pérdida de información que ello conlleva.

El presente trabajo muestra los procesos seguidos para encontrar un modelo fiable de estimación de las ventas anuales en el sector comercio de Córdoba, y cuales son las consecuencias de la falta de datos en las variables que pueden conformar el modelo.

Palabras clave:

Pequeño comercio, distribución comercial.

MODELO EXPLICATIVO DE LAS VENTAS ANUALES EN EL SECTOR PEQUEÑO COMERCIO DE CÓRDOBA. ANÁLISIS DE LA NO RESPUESTA EN LAS ENCUESTAS DE ESTUDIO DEL SECTOR.

1.- INTRODUCCIÓN

En el año 2001 la Federación Provincial del Comercio de Córdoba, encargó a la Facultad de Ciencias Económicas y Empresariales de Córdoba (ETEA) la realización del segundo estudio sobre el comercio de Córdoba

En el estudio se recorren todos los aspectos del sector presentando cómo se encuentra el mismo en la actualidad, analizando aspectos jurídicos, antigüedad en la empresa, establecimientos que posee, características de la zona comercial y el comercio y su relación con el empleo, así como describir los perfiles, tanto del empresario como del empleado, con el fin de detectar carencias que pueden ser limitativas en el desarrollo de la actividad y al mismo tiempo analizar los puntos fuertes que constituyen, sin duda la ventaja competitiva con la que cuenta el comercio.

El equipamiento comercial supone un factor importante en el desarrollo de la actividad por lo que se analizó el mismo con el fin de conocer cómo ha ido evolucionando y en qué estado se encuentra en la actualidad. Aspectos como el equipamiento informático y la distribución son carencias tradicionales en el sector constituyendo uno de los factores que limitan su desarrollo.

La actividad financiera, tipología de comercio y la publicidad y marketing completaron el análisis de los datos generales.

A continuación, el estudio analiza el tamaño de comercio y las operaciones objeto de la actividad; la compra y la venta. Desde el punto de vista de las compras se analizan factores tan importantes como medios de pago, características de los pedidos o tipología de compras y origen de los mayoristas. Desde el punto de vista de las ventas se plantea como objetivo reflejar la situación general, los medios de pago, los niveles de precios y el perfil del cliente y análisis del origen del mismo.

Cada día es más importante el conocimiento del cliente ya que donde hace unos años existía un tratamiento segmentado del mismo hoy día se habla del marketing *one to one*, es decir el conocimiento del cliente en su individualidad, esto supone que la no adaptación a los instrumentos que posibilitan este conocimiento significa perder una ventaja competitiva que hoy es determinante.

Una comparativa entre la zona de la capital y de la provincia permanente y un análisis general de las características del sector culminan los objetivos que se plantearon en el desarrollo del estudio.

El análisis del sector comercio se dividió básicamente en tres partes:

- a) Análisis de los parámetros generales que definen el sector; equipamiento, actividad financiera, tipologías de comercio y publicidad y marketing

- b) Estudio del aprovisionamiento, ya que el momento de la compra es la parte fundamental de la actividad, así entre los comerciantes se suele mantener la idea de que lo difícil es comprar y que ese momento diferencia el verdadero comerciante.
- c) Análisis de la venta como segundo gran factor a la hora de realizar la actividad.

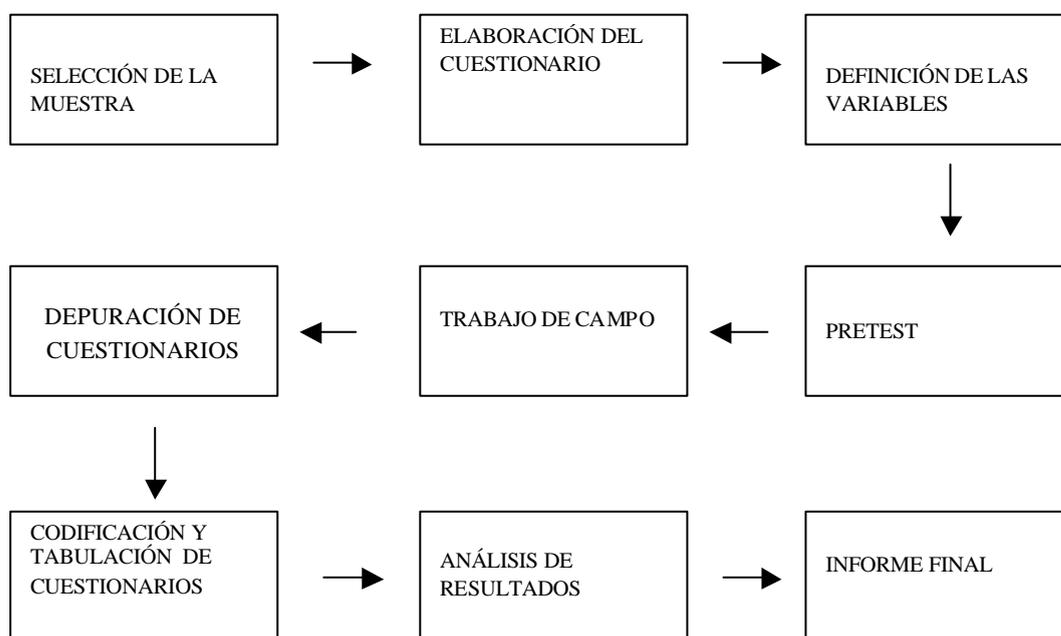
Respecto a las fuentes utilizadas, la base fundamental del trabajo ha sido obtenida de las encuestas realizadas por el grupo de investigación. Además se han realizado entrevistas a los máximos representantes del comercio en la provincia de Córdoba.

La Federación Provincial del Comercio de Córdoba encargó un primer estudio del pequeño comercio de Córdoba en 1996, que ha servido para poder realizar un análisis de la evolución en los últimos cinco años de las variables comunes tratadas en los dos estudios.

2. METODOLOGÍA

Para conseguir los objetivos que se acaban de señalar se utilizó la metodología cuyas fases figuran a continuación. Utilizando herramientas informáticas que permitieron generar las aplicaciones para el tratamiento de datos.

FIGURA 1. PROCESO SEGUIDO EN LA REALIZACIÓN DEL TRABAJO



3. GENERACIÓN DE ESTADÍSTICAS

En primer lugar, para poder efectuar una comparación de la situación del comercio cordobés en el año 2001 con el de cinco años atrás, era imprescindible la realización de una encuesta muestral del censo de empresas del comercio minorista, según las categorías derivadas de la CNAE y según criterios territoriales y demográficos.

Se trata de empresas con emplazamiento fijo, dedicadas a la venta de productos al público; se excluyeron talleres de reparación, despachos profesionales, venta ambulante y antigüedades.

Así mismo, se eligieron establecimientos con menos de 25 trabajadores, si bien el 99.2% de los comercios encuestados no excedía de 10 trabajadores.

Se tuvo en cuenta el índice de densidad comercial por lo que los pueblos seleccionados para la muestra fueron los que tenían un mayor número de establecimientos y el número de encuestas realizadas fue proporcional a los mismos.

Se seleccionaron 18 municipios (incluida la capital), en concreto aquellos que tenían más de 375 licencias, según la clasificación de actividades económicas, división 6ª, de la Cámara de Comercio del año 2000, que se corresponde básicamente a las clasificaciones 52 y 55 de la CNAE. Además de algún municipio compensatorio con menor número de licencias (véase tabla 1).

TABLA 1. INDICADORES POR MUNICIPIO

MUNICIPIO	Nº DE HABITANTES	Nº DE LICENCIAS	Nº DE ENCUESTAS
CÓRDOBA	306.248	11.120	252
AGUILAR DE LA FRONTERA	13.397	499	5
BAENA	20.057	774	20
BUJALANCE	8.204	318	1
CABRA	20.819	868	10
FERNAN NUNEZ	9.494	376	1
FUENTE PALMERA	9.695	374	6
LUCENA	35.564	1.541	45
MONTILLA	22.792	991	30
MONTORO	9.489	366	10
PALMA DEL RÍO	18.948	803	15
PEÑARROYA	13.606	493	7
POZOBLANCO	16.223	860	27
PRIEGO	22.196	964	30
PUENTE GENIL	27.918	1.140	36
RAMBLA, LA	7.229	322	8
TORRECAMPO	1.417	57	1
VILLA DEL RÍO	7.149	248	2
TOTAL	570.445	22.114	506
% SOBRE EL TOTAL DE LA PROVINCIA	75%	76%	

Fuente: Base de Datos del Comercio en la Provincia de Córdoba (en adelante BDCPCO-ETEA).

Con respecto al estudio del comercio de Córdoba realizado en 1996, se amplió en éste la muestra de municipios, que en el anterior se reducía a cuatro. Los municipios seleccionados son representativos, no sólo por el número de licencias comerciales, sino también porque aglutinan, junto con el de Córdoba capital, el 75% de la población cordobesa .

La capital cordobesa se dividió en 12 zonas cuya representación, ordenada de mayor a menor, es la siguiente: Centro, La Viñuela, Ciudad Jardín, Santa Rosa/Valdeolleros, Levante, Sector Sur, Fátima, Poniente/Miralbaida, Centro histórico, Virgen del Mar/Eroski, Trassierra, San Pablo/San Lorenzo.

4. DISEÑO Y APLICACIÓN DEL CUESTIONARIO

Es de sobra conocido la importancia que tiene, de cara a la efectividad de una encuesta, el diseño del cuestionario; por este motivo se cuidó su diseño de manera que facilitara su manejabilidad y la tarea de los encuestadores, al tiempo que se buscó una ordenación y sistematización de las preguntas buscando no disuadir al entrevistado.

Además, se tuvo en cuenta la necesidad posterior de poder cruzar las preguntas para optimizar los resultados.

Asimismo, se buscó una redacción fácil y comprensible tanto para el entrevistador como para el entrevistado.

Se buscó la pre-codificación de las preguntas aunque con un número limitado de respuestas y se introdujeron preguntas abiertas para conocer la opinión. Las preguntas figuraban con sus códigos respectivos lo que facilitó el tratamiento informatizado posterior.

El cuestionario, cuyo modelo se recoge en el Anexo V, consta de cinco grandes bloques que son los siguientes: A) datos de la empresa o negocio; B) características de la actividad; C) análisis del aprovisionamiento; D) análisis de las ventas y E) características generales.

El número de variables incluidas fue de 373.

El nivel de confianza es del 95%.

La encuesta se realizó de manera personalizada a cada uno de los comerciantes, mediante encuestadores que previamente fueron seleccionados y preparados. La encuestación tuvo lugar durante el segundo trimestre de 2001.

Una vez recopilada toda la información, ésta fue procesada y analizada. Para ello, las 373 variables estudiadas fueron tratadas informáticamente mediante el paquete estadístico SPSS (versión 10.0). Con el objetivo de profundizar todo lo posible en la realidad del sector, esta información no se examinó únicamente de forma global, sino que se ha buscado la posible existencia de diferencias significativas entre los comercios de la capital y los de la provincia¹, y (para aquellas variables en las que la desagregación podía aportar valor añadido a los comentarios) entre las distintas categorías de

¹ En concreto, se distinguieron tres grupos de comercios: aquellos que sólo tienen establecimientos en la capital, los que sólo lo poseen en otras poblaciones del resto de la provincia, y los comercios que operan simultáneamente en capital y provincia.

comercio². De igual modo, se ha aprovechado la existencia de un trabajo anterior de 1996³, para estudiar la evolución del comercio cordobés en los cinco años que median entre uno y otro trabajo.

5. LOS PROBLEMAS DE MEDICIÓN EN VARIABLES ECONÓMICAS

A pesar de todo el cuidado que se puso en la preparación del cuestionario y la preparación de los encuestadores, es conocido que las investigaciones basadas en encuestas se enfrentan a las limitaciones derivadas de la fiabilidad de la información, en especial cuando en la encuesta existen preguntas relacionadas con variables económicas (remuneraciones, ingresos, beneficios, volumen de gastos,...) donde la sospecha sobre si la respuesta dada es fiable siempre está presente.

Otro aspecto del problema, tan grave o más que el anterior, proviene cuando ni siquiera obtenemos respuesta a la pregunta. Debido a estos *valores perdidos* los estudios multivariante ven reducido el número de casos, en algunas ocasiones de forma importante, lo que reduce la fiabilidad de la muestra cara a inferencia hacia la población.

Incluso en el caso más sencillo de análisis descriptivo univariante la presencia de un elevado porcentaje de valores perdidos en la variable, puede suponer sesgos importantes en los valores de los estadísticos calculados, a no ser que se asuma o se demuestre la completa aleatoriedad en la no respuesta.

La presencia de datos ausentes o valores perdidos, también se produjo en la encuesta realizada para el estudio sobre el comercio de Córdoba. En especial hubo dos variables, ventas totales anuales y gastos totales anuales donde el índice de no respuesta llegó a superar el 60% de los casos, tal y como se puede apreciar en la tabla 2, en la que las variables 00212 y 00676 hacen referencia a la cifra de ventas anuales y los gastos totales en pesetas del año 2001, respectivamente. La citada tabla ofrece información sobre el número de casos con respuesta en dichas variables, algunos estadísticos descriptivos y la existencia de datos atípicos.

TABLA 2. SALIDA SPSS. VALORES DESCRIPTIVOS DE LAS VARIABLES VENTAS TOTALES ANUALES Y GASTOS TOTALES ANUALES.

Tabla 1: descriptivos sobre valores perdidos

	N	Media	Desviaci típ	Perdidos		Nº de extremos ^a	
				Recuento	Porcentaj	Bajo	Alto
VAR002	200	3.8E+0	1206365	306	60.	0	17
VAR002	161	2.5E+0	72619243.	345	68.	0	20

a. Número de casos fuera del rango (C1 - 1.5*AIC,

² Empleando la agrupación utilizada en el propio estudio: alimentación, equipamiento de la persona, equipamiento del hogar, salud e higiene, comercio especializado, hoteles, bares y restaurantes, otros servicios, culturales y recreativos y otros.

³ El trabajo realizado para FEPCOC “Estudio del pequeño comercio de Córdoba”, 1996.

Ante este problema los autores de la presente comunicación nos planteamos, como primera opción de solución, la posibilidad de encontrar un modelo de regresión explicativo de las ventas totales del sector comercio en Córdoba, en función de otras variables cuantitativas cuyo índice de respuesta fuera más elevado por ser menos *sensibles* para el empresario encuestado.

Dado el alto índice de no respuesta en la variable cifra anual de ventas (var00212), y la elevada variabilidad de la misma (vease el valor de la desviación típica en la tabla 1), se seleccionaron aquellos casos con información sobre la variable ventas para la obtención del modelo de regresión, es decir los 200 casos con respuesta en la mencionada variable.

6. DETERMINACIÓN DEL MODELO.

6.1. BÚSQUEDA DE VARIABLES INDEPENDIENTES.

Tras realizar un análisis exploratorio de los valores de la variable cifra anual de ventas se eliminaron 7 casos atípicos de los 200 casos con información disponible sobre dicha variable. Estos casos registraban una cifra anual de ventas extremadamente reducida (inferior a 42,07 euros, o 7.000 pesetas), que parecía indicativo de que se trataba de un error en la recogida de información. Por lo tanto el análisis de regresión se efectuó sobre 193 casos con información sobre las ventas anuales⁴. La tabla 3 muestra algunos estadísticos descriptivos de dicha variable en la muestra finalmente seleccionada.

Tabla 3: estadísticos descriptivos sobre la cifra de ventas anual

	N	Mínimo	Máximo	Media	Desv. típ.
Cifra anual de ventas	193	80000	1.4E+09	3.9E+07	1.2E+08
N válido (según lista)	193				

Para seleccionar las variables independientes que deberían de formar parte del modelo de regresión lineal que tiene por variable dependiente la cifra anual de ventas del sector pequeño comercio en Córdoba, se calcularon las correlaciones lineales bivariadas de cada variable del estudio con la variable ventas. La tabla 4 muestra información sobre los casos disponibles de cada variable, el número y porcentaje de valores perdidos, el coeficiente de correlación lineal con la variable ventas y el número de casos con información válida en cada variable y la variable ventas simultáneamente.

Las variables han sido agrupadas en variables descriptivas sobre la empresa, sobre el propietario de la misma, sobre las características físicas del establecimiento, sobre la disponibilidad de medios automatizados para la gestión, sobre la publicidad, sobre el origen de los productos (fabricación propia o ajena) y sobre ventas y clientes.

Como se deduce de la tabla 4, en la mayoría de las variables existen casos perdidos por la no respuesta, incluso en variables cuya respuesta es tan poco comprometida como las características descriptivas del

⁴ Aunque existen algunos casos con valores muy elevados de la cifra anual de ventas en comparación con la media, algunos casos, no se optó por su eliminación pues en los mencionados casos otras variables indicadoras del volumen de negocio, como la cifra de gastos, número de personal, etc. también registraban valores elevados, por lo que no se deducía un error de registro de información.

propietario (sexo, edad, experiencia, etc.). Las variables agrupadas en los conceptos sobre origen de los productos y la gestión de compras son las que ofrecen mayor grado de no respuesta (superando el 50%).

Atendiendo a los valores de los coeficientes de correlación lineal, así como el grado de significación bilateral para un nivel de significación del 5%, podemos observar que en todos los grupos de variables existe al menos una variable que correlaciona linealmente con la cifra de ventas, si bien la variable que ofrece mayor correlación es la variable gasto en primer lugar, seguida de otras como el número de empleados no familiares a tiempo completo, el número de ordenadores para la gestión, superficie de almacén y venta, número de entidades financieras, publicidad por compromiso y precio medio de cada venta.

TABLA 4: ANÁLISIS DE CORRELACIONES BIVARIADAS Y DISPONIBILIDAD DE DATOS

Variable	N	Casos Perdidos	% pérdida	Correlación lineal	Sig. bilateral	N validos con ventas
Descriptivas empresa						
Forma jurídica	200	0	0	.120	.096	193
Año constitución	199	1	0.5	.007	.920	192
% negocio act. Principal	196	4	2	-.035	.634	189
Nº establecimientos Córdoba capital	200	0	0	-.085	.242	193
Nº establecimientos provincia	200	0	0	.123	.088	193
Nº total establecimientos	200	0	0	-.017	.817	193
Nº total empleados	195	5	2.5	0.453**	0.000	188
Nº empl. familiares a.t. completo.	195	5	2.5	-.015	.839	188
Nº empl. Familiares a t. parcial	195	5	2.5	-.014	.845	188
Nº empl. no familiares a t. completo	195	5	2.5	.481**	.000	188
Nº emp no familiares a t. parcial	195	5	2.5	.247**	.001	188
Horas/semana trabajo a t. completo	173	27	13.5	.005	.948	166
Horas/semana trabajo a t. Parcial	178	22	11	.281**	.000	171
Régimen tenencia establecimiento	198	2	1	-.006	.938	191
Año apertura	178	22	11	-.065	.396	171
Pertenece a asociación	200	0	0.0	-.150*	.038	193
Cuota asociación	78	122	61.0	.354**	.002	77
Descriptivos del propietario						
Edad propietario	172	28	14	-.046	.553	165
Sexo propietario	178	22	11	.093	.228	171
Nivel estudios	170	30	15	-.137	.080	164
Años dedicación	100	100	50	-.024	.813	97
Porcentaje depend. activ. comercial	175	25	12.5	.024	.757	168
Importancia comercial zona	200	0	0	.053	.462	193
Formación empresarial	159	41	20.5	-.230**	.004	152
Descriptivos del establecimiento						
m ² superficie venta	200	0	0	.781**	.000	193
m ² superficie almacén	165	35	17.5	.791**	.000	158
m ² superficie oficina	36	164	82	.360*	.036	34
m. mostrador	190	10	5	.065	.382	183
m. escaparate	187	13	6.5	.106	.154	181
m. fachada	194	6	3	.301**	.000	187
Nº puertas	200	0	0	.457**	.000	193
Nº plantas	172	28	14	.015	.844	167
Climatización	196	4	2	.077	.292	189
Musica ambiental	198	2	1	.053	.467	191
Equipación tecnología, informatización y medios distribución						
Nº cajas registradoras	140	60	30	.273**	.001	133
Nº terminales tpv	73	127	63.5	.071	.555	72
Nº ordenadores para la gestión	200	0	0.0	.504**	.000	193
Año informatización	55	141	72.5	-.297*	.029	54
Utiliza base datos	161	39	19.5	-.371**	.000	154
Gestiona listas personal inform	39	161	80.5	-.425**	.000	147

Variable	N	Casos Perdidos	% pérdida	Correlación lineal	Sig. bilateral	N validos con ventas
Gestiona listas inform proveedores	186	14	7.0	-.256**	.001	179
Gestiona listas inform productos	176	24	12.0	-.277**	.000	169
Consulta páginas web	149	51	25.5	-.304**	.000	143
Correo electrónico privado	151	49	24.5	-.464**	.000	145
Página Web informativa	142	58	29.0	-.433**	.000	136
Posee vehículos la empresa	200	0	0.0	-.193**	.007	193
Nº coches de empresa	147	53	26.5	.273**	.001	140
Antigüedad media en años coches empresa	22	178	89	-.240	.282	22
Nº furgonetas empresa	65	135	67.5	.165	.189	65
Antigüedad media en años furgonetas	61	139	69.5	-.164	.207	61
Nº camiones	2	198	99	-1.000**	.	2
Antigüedad media en años camiones	2	198	99	-1.000**	.	2
Nº de vehículos propietario	89	111	55	-.033	.770	83
Antigüedad vehículos propios	85	115	57.5	-.027	.813	79

TABLA 4: ANÁLISIS DE CORRELACIONES BIVARIADAS Y DISPONIBILIDAD DE DATOS (CONTINUACIÓN)

Variable	N	Casos Perdidos	% pérdida	Correlación lineal	Sig. bilateral	N validos con ventas
Comunicación y publicidad						
Hace publicidad	200	0	0.0	.190**	.008	193
Comunicación y publicidad						
Hace publicidad	200	0	0.0	.190**	.008	193
Buzones	200	0	0.0	.063	.387	193
Coches	200	0	0.0	.009	.904	193
Mailing	200	0	0.0	.177*	.014	193
Prensa	200	0	0.0	.201*	.005	193
Radio	200	0	0.0	.115	.111	193
Tv	200	0	0.0	.055	.451	193
Compromisos	200	0	0.0	.455**	.000	193
Web no propia	200	0	0.0	-.003	.965	193
Páginas amarillas	200	0	0.0	.131	.069	193
Otros medios publicitarios	200	0	0.0	-.019	.797	193
Posee catálogos	199	1	0.5	-.289**	.000	192
Origen productos						
% productos fabricación propia	150	50	25.0	.056	.503	143
% productos fabricación ajena en Córdoba	150	50	25.0	-.106	.209	143
% productos fabricación ajena en Andalucía	150	50	25.0	-.106	.209	143
% productos fabricación ajena en resto España	150	50	25.0	-.046	.589	143
% productos fabricación ajena en Extranjero	150	50	25.0	.273**	.001	143
Compras y aprovisionamiento						
Nº proveedores que acuden establec.	171	29	14.5	.354**	.000	164
Nº medio artículos por pedido	170	30	15.0	.108	.168	163
Importe medio pedido	167	33	16.5	.094	.235	160
Compras						
Nº proveedores in situ	81	119	59.5	-.031	.791	75
Nº medio artículos pedido in situ	94	106	53.0	.273*	.010	88
Importe medio pedido in situ	92	108	54.0	.130	.231	86
Nº proveedores por pedido	71	129	64.5	.117	.354	65
Nº medio artículos pedido	78	122	61.0	.312**	.008	72
Importe medio pedido	78	122	61.0	.167	.160	72
Precio medio compra	195	5	2.5	.198**	.006	188
Nº entidades financieras	187	13	6.5	.340**	.000	182
Ventas y clientes						
Principal sistema de venta	200	0	0.0	.198**	.006	188
Ofrece pedido por teléfono	200	0	0.0	-.210**	.003	193
Ofrece pago aplazado sin interés	200	0	0.0	-.188**	.009	193
Personaliza el producto	200	0	0.0	-.181*	.012	193
Precio medio venta	183	17	8.5	.478**	.000	176
Cifra anual de ventas	193	7	3.5	.1		193

Variable	N	Casos Perdidos	% pérdida	Correlación lineal	Sig. bilateral	N validos con ventas
Margen comercial medio	138	62	31.0	-.112	.200	132
Gastos totales	159	41	20.5	.958**	.000	152
% clientes hombres	178	22	11.0	.283**	.000	172
% clientes mujeres	178	22	11.0	-.318**	.000	172
Edad media clientes	177	23	11.5	-.035	.647	170
Poder adquisitivo medio	198	2	1.0	.105	.147	191
% clientes zona	196	4	2.0	-.187**	.009	191
% clientes otras zonas	196	4	2.0	.016	.822	191
% clientes provincia	196	4	2.0	.109	.135	191
% clientes otras provincias	196	4	2.0	.389**	.000	191
Pertenece a asociación	200	0	0.0	-.150*	.038	193
Cuota asociación	78	122	61.0	.354**	.002	77
Importancia comercial zona	200	0	0	.053	.462	193

6.2. MODELO DE REGRESIÓN

Se utilizó el programa estadístico SPSS versión 10 bajo entorno Windows para efectuar el análisis de regresión lineal múltiple. Se introdujeron todas las variables cuantitativas del estudio y se empleó el método de selección progresivo para seleccionar las mejores variables independientes del modelo.

Como era de esperar la variable Gastos totales, la que ofrecía mayor grado de correlación con la cifra de ventas anual es la primera que se introduce en el modelo, seguida de la variable precio medio de cada venta (que expresa el precio medio de venta de los artículos), y número de ordenadores para la gestión.

La salida de Spss del modelo anteriormente descrito se ofrece a continuación. Como podemos ver el coeficiente de determinación lineal (igual al 94,4%) es bastante aceptable y tanto el modelo como los coeficientes de regresión son significativos. Los estadísticos sobre multicolinealidad muestran unos valores de tolerancia aceptables.

TABLA 5. PRIMER MODELO DE REGRESIÓN

VARIABLES INTRODUCIDAS/ELIMINADAS^{a,b}

Modelo	VARIABLES INTRODUCIDAS	VARIABLES ELIMINADAS	Método
1	Gastos totales	.	Por pasos (criterio: Probabilidad de F para entrar \leq .050, Probabilidad de F para salir \geq .100).
2	precio medio de cada venta	.	Por pasos (criterio: Probabilidad de F para entrar \leq .050, Probabilidad de F para salir \geq .100).
3	Nº ordenadores para la gestión	.	Por pasos (criterio: Probabilidad de F para entrar \leq .050, Probabilidad de F para salir \geq .100).

a. Variable dependiente: Cifra anual de ventas

b. Regresión lineal a través del origen

TABLA 6. RESUMEN DEL PRIMER MODELO DE REGRESIÓN

Resumen del modelo ^{e,f}

Modelo	R	R cuadrado ^a	R cuadrado corregido	Error típ. de la estimación	Cambiar los estadísticos				
					Cambio en R cuadrado	Cambio en F	gl1	gl2	Sig. del cambio en F
1	.959 ^o	.920	.919	21985431.62	.920	1620.433	1	141	.000
2	.968 ^c	.938	.937	19492335.01	.018	39.375	1	140	.000
3	.972 ^d	.945	.944	18374578.53	.007	18.551	1	139	.000

a. Para la regresión a través del origen (el modelo sin intersección), R cuadrado mide la proporción de la variabilidad en la variable dependiente acerca del origen explicado por la regresión. No es posible comparar esto con R cuadrado para los modelos que incluyen una intersección.

b. Variables predictoras: Gastos totales

c. Variables predictoras: Gastos totales , precio medio de cada venta

d. Variables predictoras: Gastos totales , precio medio de cada venta, N° ordenadores para la gestión

e. Variable dependiente: Cifra anual de ventas

f. Regresión lineal a través del origen

TABLA 7. ANÁLISIS ANOVA SOBRE EL PRIMER MODELO DE REGRESIÓN

ANOVA ^{e,f}

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	7.833E+17	1	7.83E+17	1620.433	.000 ^a
	Residual	6.815E+16	141	4.83E+14		
	Total	8.514E+17 ^b	142			
2	Regresión	7.982E+17	2	3.99E+17	1050.414	.000 ^c
	Residual	5.319E+16	140	3.80E+14		
	Total	8.514E+17 ^b	142			
3	Regresión	8.045E+17	3	2.68E+17	794.249	.000 ^d
	Residual	4.693E+16	139	3.38E+14		
	Total	8.514E+17 ^b	142			

a. Variables predictoras: Gastos totales

b. Esta suma total de cuadrados no se ha corregido para la constante porque la constante es cero para la regresión a través del origen.

c. Variables predictoras: Gastos totales , precio medio de cada venta

d. Variables predictoras: Gastos totales , precio medio de cada venta, N° ordenadores para la gestión

e. Variable dependiente: Cifra anual de ventas

f. Regresión lineal a través del origen

TABLA 8. ANÁLISIS DE COEFICIENTES SOBRE EL PRIMER MODELO DE REGRESIÓN

Coefficientes ^{a,b}

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.	Estadísticos de colinealidad	
		B	Error típ.	Beta			Tolerancia	FIV
1	Gastos totales	1.137	.028	.959	40.255	.000	1.000	1.000
2	Gastos totales	1.070	.027	.903	39.313	.000	.846	1.181
	precio medio de cada venta	458.211	73.023	.144	6.275	.000	.846	1.181
3	Gastos totales	1.013	.029	.854	35.012	.000	.666	1.501
	precio medio de cada venta	349.902	73.285	.110	4.775	.000	.747	1.339
	Nº ordenadores para la gestión	8095320	1879538	.110	4.307	.000	.605	1.653

a. Variable dependiente: Cifra anual de ventas

b. Regresión lineal a través del origen

TABLA 9. PROCESO POR PASOS DEL PRIMER MODELO DE REGRESIÓN

Variables excluidas^{d,e}

Modelo		Beta dentro	t	Sig.	Correlación parcial	Estadísticos de colinealidad		
						Tolerancia	FIV	Tolerancia mínima
1	m2 superficie de almacén	.039 ^a	1.477	.142	.124	.788	1.269	.788
	Nº ordenadores para la gestión	.152 ^a	5.888	.000	.446	.686	1.459	.686
	Compromisos	.071 ^a	2.891	.004	.237	.892	1.120	.892
	precio medio de cada venta	.144 ^a	6.275	.000	.469	.846	1.181	.846
	% de clientes de otras provincias	.011 ^a	.452	.652	.038	.995	1.005	.995
2	m2 superficie de almacén	.018 ^b	.728	.468	.062	.771	1.298	.718
	Nº ordenadores para la gestión	.110 ^b	4.307	.000	.343	.605	1.653	.605
	Compromisos	.033 ^b	1.427	.156	.120	.818	1.223	.775
	% de clientes de otras provincias	.007 ^b	.337	.736	.029	.995	1.005	.844
3	m2 superficie de almacén	-.005 ^c	-.220	.826	-.019	.730	1.370	.573
	Compromisos	.015 ^c	.673	.502	.057	.786	1.272	.581
	% de clientes de otras provincias	.001 ^c	.064	.949	.005	.990	1.010	.602

a. Variables predictoras en el modelo: Gastos totales

b. Variables predictoras en el modelo: Gastos totales , precio medio de cada venta

c. Variables predictoras en el modelo: Gastos totales , precio medio de cada venta, Nº ordenadores para la gestión

d. Variable dependiente: Cifra anual de ventas

e. Regresión lineal a través del origen

Como la principal variable explicativa del modelo anterior es Gastos totales, y esta es precisamente una de las variables con mayor ausencia de respuesta en las encuestas, se buscó un modelo alternativo al anterior que explique la cifra de ventas anual en función de otras variables cuyo grado de respuesta sea más elevado, así como la fiabilidad de la misma. Al prescindir de la variable explicativa Gastos totales, el mejor modelo lineal se ofrece en la siguiente salida:

Modelo alternativo:

TABLA 10. MODELO ALTERNATIVO DE REGRESIÓN

VARIABLES INTRODUCIDAS/ELIMINADAS^a

Modelo	VARIABLES INTRODUCIDAS	VARIABLES ELIMINADAS	Método
1	m ² superficie de almacén	.	Por pasos (criterio: Prob. de F para entrar <= .050, Prob. de F para salir >= .100).
2	precio medio de cada venta	.	Por pasos (criterio: Prob. de F para entrar <= .050, Prob. de F para salir >= .100).
3	Número de empleados no familiares a tiempo completo	.	Por pasos (criterio: Prob. de F para entrar <= .050, Prob. de F para salir >= .100).
4	numero de entidades financieras	.	Por pasos (criterio: Prob. de F para entrar <= .050, Prob. de F para salir >= .100).
5	Nº ordenadores para la gestión	.	Por pasos (criterio: Prob. de F para entrar <= .050, Prob. de F para salir >= .100).

a. Variable dependiente: Cifra anual de ventas

b. Regresión lineal a través del origen

TABLA 11. RESUMEN DEL MODELO ALTERNATIVO DE REGRESIÓN

Resumen del modelo

Modelo	R	R cuadrado ^a	R cuadrado corregida	Error típ. de la estimación
1	.847 ^b	.718	.716	74319307.46
2	.924 ^c	.854	.852	53606831.21
3	.935 ^d	.874	.871	50029270.80
4	.938 ^e	.880	.877	48958873.06
5	.942 ^f	.886	.882	47890121.53

- a. Para la regresión a través del origen (el modelo sin término de intersección), R cuadrado mide la proporción de la variabilidad de la variable dependiente explicado por la regresión a través del origen. NO SE PUEDE comparar lo anterior con la R cuadrado para los modelos que incluyen una intersección.
- b. Variables predictoras: m2 superficie de almacén
- c. Variables predictoras: m2 superficie de almacén, precio medio de cada venta
- d. Variables predictoras: m2 superficie de almacén, precio medio de cada venta, Número de empleados no familiares a tiempo completo
- e. Variables predictoras: m2 superficie de almacén, precio medio de cada venta, Número de empleados no familiares a tiempo completo, numero de entidades financieras
- f. Variables predictoras: m2 superficie de almacén, precio medio de cada venta, Número de empleados no familiares a tiempo completo, numero de entidades financieras, N° ordenadores para la gestión

TABLA 12. ANÁLISIS ANOVA SOBRE EL MODELO ALTERNATIVO DE REGRESIÓN

ANOVA^{g,h}

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	1.885E+18	1	1.89E+18	341.285	.000 ^a
	Residual	7.401E+17	134	5.52E+15		
	Total	2.625E+18 ^b	135			
2	Regresión	2.243E+18	2	1.12E+18	390.259	.000 ^c
	Residual	3.822E+17	133	2.87E+15		
	Total	2.625E+18 ^b	135			
3	Regresión	2.295E+18	3	7.65E+17	305.613	.000 ^d
	Residual	3.304E+17	132	2.50E+15		
	Total	2.625E+18 ^b	135			
4	Regresión	2.311E+18	4	5.78E+17	241.051	.000 ^e
	Residual	3.140E+17	131	2.40E+15		
	Total	2.625E+18 ^b	135			
5	Regresión	2.327E+18	5	4.65E+17	202.926	.000 ^f
	Residual	2.982E+17	130	2.29E+15		
	Total	2.625E+18 ^b	135			

- a. Variables predictoras: m2 superficie de almacén
- b. Esta suma de cuadrados total no se ha corregido para la constante porque la constante es cero para la regresión a través del origen.
- c. Variables predictoras: m2 superficie de almacén, precio medio de cada venta
- d. Variables predictoras: m2 superficie de almacén, precio medio de cada venta, Número de empleados no familiares a tiempo completo
- e. Variables predictoras: m2 superficie de almacén, precio medio de cada venta, Número de empleados no familiares a tiempo completo, numero de entidades financieras
- f. Variables predictoras: m2 superficie de almacén, precio medio de cada venta, Número de empleados no familiares a tiempo completo, numero de entidades financieras, N° ordenadores para la gestión
- g. Variable dependiente: Cifra anual de ventas
- h. Regresión lineal a través del origen

TABLA 13. ANÁLISIS DE COEFICIENTES SOBRE EL MODELO ALTERNATIVO DE REGRESIÓN

Coeficientes^{a,b}

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.
		B	Error típ.	Beta		
1	m2 superficie de almacén	137052.4	7418.705	.847	18.474	.000
2	m2 superficie de almacén	96205.172	6483.098	.595	14.839	.000
	precio medio de cada venta	3899.621	349.417	.447	11.160	.000
3	m2 superficie de almacén	92650.676	6100.662	.573	15.187	.000
	precio medio de cada venta	3205.449	360.024	.368	8.903	.000
	Número de empleados no familiares a tiempo completo	2.0E+07	4399166	.169	4.550	.000
4	m2 superficie de almacén	90757.225	6013.905	.561	15.091	.000
	precio medio de cada venta	3615.488	385.654	.415	9.375	.000
	Número de empleados no familiares a tiempo completo	2.5E+07	4640716	.207	5.289	.000
	numero de entidades financieras	-8305385	3176809	-.105	-2.614	.010
5	m2 superficie de almacén	91348.796	5886.926	.565	15.517	.000
	precio medio de cada venta	3332.297	392.312	.382	8.494	.000
	Número de empleados no familiares a tiempo completo	1.8E+07	5202447	.151	3.434	.001
	numero de entidades financieras	-9433708	3136956	-.119	-3.007	.003
	Nº ordenadores para la gestión	1.5E+07	5812152	.117	2.629	.010

a. Variable dependiente: Cifra anual de ventas

b. Regresión lineal a través del origen

TABLA 14. PROCESO POR PASOS DEL MODELO ALTERNATIVO DE REGRESIÓN

Variables excluidas^{e,f}

Modelo	Beta dentro	t	Sig.	Correlación parcial	Estadísticos de colinealidad	
					Tolerancia	
1	Número de empleados no familiares a tiempo completo	.309 ^a	7.291	.000	.534	.843
	numero de entidades financieras	.187 ^a	4.139	.000	.338	.918
	Nº ordenadores para la gestión	.324 ^a	8.015	.000	.571	.876
	precio medio de cada venta	.447 ^a	11.160	.000	.695	.681
2	Número de empleados no familiares a tiempo completo	.169 ^b	4.550	.000	.368	.692
	numero de entidades financieras	-.026 ^b	-.628	.531	-.055	.657
	Nº ordenadores para la gestión	.165 ^b	4.244	.000	.347	.645
3	numero de entidades financieras	-.105 ^c	-2.614	.010	-.223	.565
	Nº ordenadores para la gestión	.098 ^c	2.173	.032	.187	.452
4	Nº ordenadores para la gestión	.117 ^d	2.629	.010	.225	.444

a. Variables predictoras en el modelo: m2 superficie de almacén

b. Variables predictoras en el modelo: m2 superficie de almacén, precio medio de cada venta

c. Variables predictoras en el modelo: m2 superficie de almacén, precio medio de cada venta, Número de empleados no familiares a tiempo completo

d. Variables predictoras en el modelo: m2 superficie de almacén, precio medio de cada venta, Número de empleados no familiares a tiempo completo, numero de entidades financieras

e. Variable dependiente: Cifra anual de ventas

f. Regresión lineal a través del origen

Como se puede observar en el modelo alternativo la bondad de ajuste es más baja (del 88,2%), aunque también es válido en términos de significación global y de la de sus coeficientes.

7. Conclusiones

A modo de resumen final se pueden extraer las siguientes conclusiones:

- Cuando se realizan investigaciones en la que se incluyen variables relacionados con gastos e ingresos de familias o empresas es habitual encontrar un número más o menos elevado de no respuesta. En el caso del último estudio sobre el comercio de Córdoba, se llega a índices del 60% de valores perdidos en las variables ingresos totales anuales y gastos totales anuales. Este elevado número de valores perdidos suponen un problema para el investigador, puesto que reduce la significación de la muestra, y en el caso de que se demuestre que los valores perdidos mantienen algún tipo patrón de comportamiento frente a otras variables, invalida los análisis estadísticos por el sesgo que se está introduciendo.
- Siendo conscientes de este problema los autores del presente trabajo nos hemos planteado la posibilidad de encontrar un modelo de regresión explicativo de las ventas totales en función de otras variables que nos permita trabajar aquellos casos de interés en los que nos falte dicha información.
- Dentro del análisis del resto de variables que conformaban el estudio sobre el comercio de Córdoba, se encontraron un elevado número de variables influyentes en el volumen de ventas, que como es natural son variables relacionadas con otros indicadores significativos del volumen de actividad de un negocio.
- En la búsqueda de un modelo de regresión explicativo se han utilizado dos modelos distintos. El primer modelo utiliza tres variables: los gastos totales, el precio medio de cada venta y el número de ordenadores utilizados para la gestión, que alcanza un coeficiente R^2 del 0,945. Sin embargo el hecho de que la principal variable explicativa del modelo sean los gastos totales, y esta variable a su vez sea una de las que más valores perdidos tiene nos obliga a buscar otro modelo que con un valor de R^2 aceptable incluya variables con menores porcentajes de valores perdidos. Dicho modelo lo encontramos al usar las variables metros cuadrados de superficie de almacén, precio medio de cada venta, nº de empleados a tiempo completo no familiares, nº de entidades financieras con la que se trabaja y nº de ordenadores usados en la gestión del negocio. Este modelo alcanza un R^2 de 0,886, que aún siendo inferior al anterior es un valor elevado, y las variables usadas alcanzan siempre valores mayores de respuesta que las del primer modelo, por lo que puede hacerlo preferible.
- El procedimiento seguido y descrito en la presente comunicación consiste un primer paso del análisis que de las posibles soluciones a la no respuesta pueden darse y nos planteamos seguir dicho análisis a través de las siguientes líneas de trabajo:
 - Análisis de validez de diferentes algoritmos de sustitución de valores perdidos:
 - Método EM
 - Método de Regresión (a través de los módulos de software que para análisis de valores perdidos incluyen los paquetes estadísticos)
 - Método de imputación múltiple.
 - Análisis de otros modelos clásicos de regresión no lineales.
 - Análisis de otros modelos basados en redes neuronales.
 - Análisis de los modelos introduciendo variables no métricas, en especial variables de opinión o de perfil del empresario.