

**CONGRESO DE LA ASOCIACIÓN ESPAÑOLA DE CIENCIA REGIONAL
XXX REUNIÓN DE ESTUDIOS REGIONALES
"LA POLÍTICA REGIONAL EN LA ENCRUCIJADA"
Barcelona, 18-19 de noviembre de 2004**

***"El Comercio Electrónico en la Región de Murcia: Modelo de Regresión
Logística aplicado a las pymes."***

Bernal García, Juan Jesús

Catedrático de Economía Aplicada. juanjesus.bernal@upct.es

Martínez María-Dolores, Soledad María

Profesora Ayudante de Universidad. soledad.martinez@upct.es

Sánchez García, Juan Francisco

Profesor Ayudante de Universidad. jf.sanchez@upct.es

Dpto. de Métodos Cuantitativos e Informáticos.

Universidad Politécnica de Cartagena

Resumen

Mediante una encuesta efectuada a una muestra de empresas de la Región de Murcia se han obtenido datos reales y fiables que nos indican cuáles son las variables que pueden tener más relevancia para el desarrollo del comercio electrónico en esta región. A su vez, estos datos indagan acerca de qué entienden los empresarios consultados por realizar comercio electrónico en su empresa, y a partir de estos se desarrolla un modelo mediante la utilización del análisis de regresión logística, que nos permite caracterizar a las empresas que deciden hacer comercio electrónico en base a determinadas variables que se consideran pueden ser relevantes bien como objetivos a alcanzar o bien como barreras básicas a solucionar, detectando a su vez si las características intrínsecas de las empresas murcianas también resultan influyentes. Este modelo se ha implementado en una herramienta de hoja de cálculo para facilitar su utilización posterior tanto por la Administración Pública como por las propias empresas murcianas.

Palabras clave: e-commerce, modelos de regresión logística, probabilidad, pymes.

Área Temática: Técnicas de análisis regional.

Introducción

Al iniciar esta investigación pretendíamos aproximarnos a la situación que las empresas murcianas tiene respecto al Comercio Electrónico, obteniendo datos reales y fiables mediante la utilización en primer lugar de un panel de expertos que en un primer análisis nos hizo comprobar si se podía realizar una investigación coherente en torno a este tema, y en un segundo término utilizando una técnica de encuestas obtuvimos un espacio muestral de 450 empresas de la Región, diversificadas por tamaños y sectores, que consideramos que son una fiel representación de la estructura empresarial de Murcia.

Es importante, debido al tema tratado, que el periodo de tiempo en el que fueran

recogidos los datos fuese el mínimo posible, por lo que los mismos no presentan grandes sesgos debidos

Cálculo del Tamaño Muestral

1. No se conoce el tamaño de la población

Probabilidad	95%		
Alfa (a)	0,05		
Nivel confianza (Z)	97,50%	1,96	(Ajustado a la normal)
Variación positiva (p)	50%		
Variación negativa (q)	50%		
Error o precisión (E)	4,00%	Error o precisión (E)	4,62%

Tamaño muestral población no conocida (n)	600	empresas	450
---	------------	----------	------------

2. Si se conoce el tamaño de la población

Tamaño población (N)	100.000	empresas	
Tamaño muestral población si es conocida (N)	597	empresas	448
Proporción presente (P)	0,00596646		
Tolerancia de la muestra	0,006179429		
Intervalo de confianza (p) = (P ± tolerancia de la muestra)	0,012145889 -0,000212969	superior inferior	↕ 1.214 ↕ - 22

Tabla 1 Cálculo del Tamaño Muestral

a la temporalidad eliminando de esta forma los errores que se pudieran reflejar por grandes cambios en el contexto entre los encuestados.

Para estimar cuál sería el tamaño óptimo a alcanzar de la muestra obteniendo una estimación fiable de la realidad empresarial en la Región de Murcia con referencia a estos temas, realizamos los cálculos que se presentan en la *Tabla 1*. Si no se conoce el tamaño de la población, como ocurría en nuestro caso, para un nivel de confianza del 97,50% ajustado a una curva normal, con un 95% de probabilidad y asumiendo un error del 4% sería necesario un tamaño muestral de 600 empresas. Si asumimos un error del 4,62% obtendríamos un

tamaño muestral para un tamaño de población no conocido de 450 empresas. También se muestran los cálculos realizados acerca de la dimensión muestral idónea en función de que el parámetro de tamaño de la población fuese conocido asumiendo el supuesto, ya que esta cifra no la conocíamos en realidad, que el número total de empresas de la Región fuera de unas 100.000, con lo que obteníamos una muestra de 597 para el caso de que fuera el 4% el error soportado, y de 448 en el caso de que fuese del 4,62%.

El envío del cuestionario se realizó fundamentalmente a través de tres vías: utilizando las cuentas de correo de una base de datos de empresas de la Región

Total Encuestas Obtenidas	537
Total Encuestas Válidas	450
× Web	152
× Fax	21
× Correo Electrónico	219
× Entrevista	58

Tabla 2

de Murcia facilitada por el Instituto de Fomento (INFO), a través de la Confederación Comarcal de Organizaciones Empresariales de Cartagena (COEC) que la envió a sus entidades miembro mediante un encarte en la revista de Noviembre de 2003, y mediante la Confederación Regional de Organizaciones Empresariales de Murcia (CROEM) que reenvió un correo electrónico a todas sus entidades miembro con el archivo adjunto del cuestionario. Se obtuvieron un total de 537 respuestas de distintas empresas de la Región de las que finalmente resultaron válidas¹ un total de 450, tal y como se puede observar en la *Tabla 2*.

El cuestionario se compone en su mayoría por variables de tipo categórico, por lo que se realizaron los correspondientes análisis de bondad del ajuste, independencia y fiabilidad de las respuestas obtenidas, no presentando ninguna de las variables ajuste a ninguna distribución conocida, una vez realizadas las pruebas de *Chi-cuadrado de Pearson*, *Binomial* o de *Kolmogorov-Smirnov* en función del tipo de variable que estuviéramos evaluando. La fiabilidad para contrastar la capacidad de la escala de medir de forma consistente y precisa, para cada conjunto de variables, la característica que se pretendía medir en cada una de las preguntas la realizamos mediante el

¹ El motivo principal para desechar algunas de las respuestas fueron causas de forma, ya que se cometieron errores al reenviar el archivo o porque éste no aparecía bien cumplimentado, debido casi siempre a que el encuestado no había guardado los cambios efectuados en el mismo. La última encuesta válida obtenida se recibió el 29 de Febrero de 2004.

método del parámetro alfa o de Cronbach, pudiendo determinar así la de variabilidad que depende de la variable y qué parte de esta variabilidad pertenece a los errores que se desprenden en el proceso de medición, obteniendo una aproximación muy buena en las escalas planteadas².

Nivel de implantación del Comercio electrónico en Murcia

En esta comunicación, vamos solamente a incidir en las cuestiones planteadas en el cuestionario acerca del nivel de implantación del comercio electrónico en la Región de Murcia y en el análisis de regresión logística realizado para conocer qué variables influyen en la probabilidad de que una empresa realice o no actividades de este tipo.

Destacar que el 33,33% de las empresas consultadas afirma realizar alguna actividad de este tipo, aunque realizando un desglose mayor, podemos concluir que e-commerce real solamente lo efectúan el 52,38% de las que contestaron positivamente. Por lo tanto el número de empresas que implican de una forma completa su forma de hacer negocio por la red es bastante bajo, tanto en transacciones destinadas al consumidor final como entre empresas. Se ha podido detectar que en muchas ocasiones, para una empresa, el realizar comercio electrónico supone tener solamente una presencia en la web mediante una pequeña página, o a insertar su nombre en bases de datos del sector o a utilizar el correo electrónico como forma de ponerse en contacto con los clientes. Queremos indicar, que si bien estas actividades se pueden considerar como una primera aproximación a realizar comercio electrónico, tal vez no sea suficiente, por lo que es importante que las que han demostrado un interés al menos inicial con respecto al uso de esta tecnología y que han vencido ese miedo primero a la utilización de un medio desconocido, se encuentren con incentivos para seguir incorporando paso a paso su negocio a la red.

De las empresas de la muestra consideradas grandes³ por el número de empleados con el que cuentan, comprobamos que el 44,44% ya realiza alguna actividad de comercio electrónico, sin embargo las de tamaño mediano

² Este parámetro considera muy buena la escala cuanto más próximo a uno se encuentra, y nuestra escala presentaba en todos los casos valores superiores al 0,85.

³ Grandes = Más de 250 empleados; Medianas = entre 50 y 250 empleados; Pequeñas = entre 5 y 50 empleados; Microempresas = menos de 5 empleados; Cybertrader = sólo presencia en la red.

solamente son un 36,73% de las de su categoría, correspondiendo un 31,97% y un 28,42% a las pequeñas y microempresas respectivamente.

La distribución de las empresas que afirman realizar alguna política de Comercio Electrónico según el sector al que pertenecen es la siguiente: 93 empresas son del Sector Servicios, 51 pertenecen al Sector Industrial y 6 al de la Construcción. De todas las empresas que contestaron la encuesta y que pertenecen al sector Servicios, solamente el 38,11% realiza actividades de comercio electrónico en la actualidad. Si tenemos en cuenta que éste es uno de los principales motores dentro de la Economía Murciana, y que los productos o servicios que se venden por la red son especialmente apropiados para este

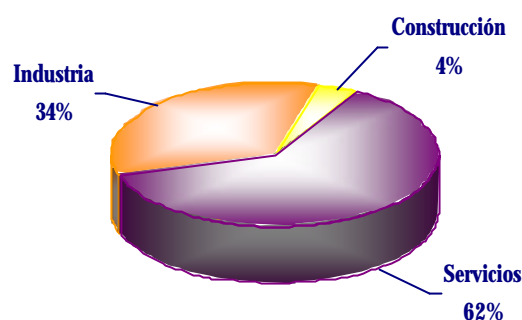


Figura 1. Sector al que pertenecen las empresas que Sí hacen Comercio Electrónico

sector, consideramos que se han de potenciar los esfuerzos para realizar acciones efectivas dentro del mismo, ya que en principio se configura como el sector más predispuesto y/o preparado para lanzarse a negociar por la Red.

Dejamos un ítem abierto, que se incluía dentro de la misma pregunta, para que en el caso de que se diera una respuesta afirmativa por parte de una empresa, ésta se justificara explicando qué tipo de acción se llevaba a cabo en la misma respecto al comercio electrónico. Esta pregunta nos reveló claramente qué es lo que la empresa entiende por hacer comercio electrónico, pudiendo contrastarse con posterioridad si se ajustaban o no las respuestas obtenidas a lo que de verdad significa. Por otro lado, es un sistema clásico en las encuestas para que no se conteste una opción u otra sin necesidad de tener que dar más explicaciones al respecto, pudiendo así exagerar los datos. Nos encontramos por tanto una amplia variedad de respuestas, que hemos agrupado de la siguiente forma:

a) Referencia en la Red (4,76%). Incluye a las empresas que indicaron que la acción con respecto al comercio electrónico que realizaban era simplemente encontrarse dadas de alta en alguna base de datos de algún organismo oficial o en listas de distribución de clientes potenciales, o que aparecieran sus datos en guías, páginas web o portales relacionados con su sector. También se han

incluido en esta categoría aquellas empresas que indicaban que tenían correo y acceso a Internet, y que estaban estudiando el realizar próximamente su propia página web.

b) Página Web (9,52%). Incluye a las empresas que nos indicaron que tenían una página web colgada ya en la Red. Visitamos dichas direcciones, y el criterio para situarlas dentro de este grupo, y no en el que a continuación mostramos, fue que simplemente eran páginas que daban información acerca de la empresa, no eran actualizadas de forma continua, y en muchas ocasiones tenían como mucho una sola página de presentación, aunque en algunas también figuraba una dirección de correo electrónico por si alguien quería ponerse en contacto. También situamos en este grupo las páginas de las empresas en la Red que se encontraban aún en construcción.

c) Catálogo-Información-Transmisión a Terceros (14,29%). En este grupo se sitúan las empresas que a través de Internet y de su página, ofertan o bien sus productos o bien sus servicios, mantienen un catálogo actualizado de los mismos, utilizan este medio para hacer publicidad, a través de la red o del correo electrónico, lo usan como medio para establecer los primeros contactos con los nuevos y potenciales clientes, facilitan presupuestos y atienden consultas, o realizan transacciones con bancos de forma online.

d) Pedidos-Venta (19,05%). En esta categoría incorporamos un paso más al proceso de transacción online, pudiendo así encontrarnos con páginas que mantienen una política comercial activa con clientes que figuran en sus bases de datos y en donde se permite la recepción de pedidos a través de la web. Se podrían considerar por tanto como tiendas virtuales que tienen un catálogo y que permiten efectuar pedidos ejecutando de esta forma la venta, pero sólo hasta este paso, ya que no permiten pagar a través de la red ni hacer un seguimiento posterior, ni cancelar dicho pedido o facilitar un servicio postventa.

e) B2C (23,81%). Este grupo es el que finalmente sí que tiene instaurada una política completa de e-commerce en su empresa hacia al consumidor final, incorporando así todo tipo de servicios y facilidades para completar el proceso de transacción comercial.

f) B2B (28,57%). Son empresas que especificaron que utilizaban la tecnología para comunicarse mediante plataformas business to business, es decir

realizando comercio electrónico entre empresas, con sus proveedores o con otras empresas clientes. Muchas ellas también realizan B2C, es decir que podrían estar incluidas en el apartado anterior, pero al ser más amplio su campo de actuación mediante el uso de esta tecnología, hemos considerado importante el realizar esta separación. En la *Figura 2* se muestran gráficamente los resultados obtenidos.

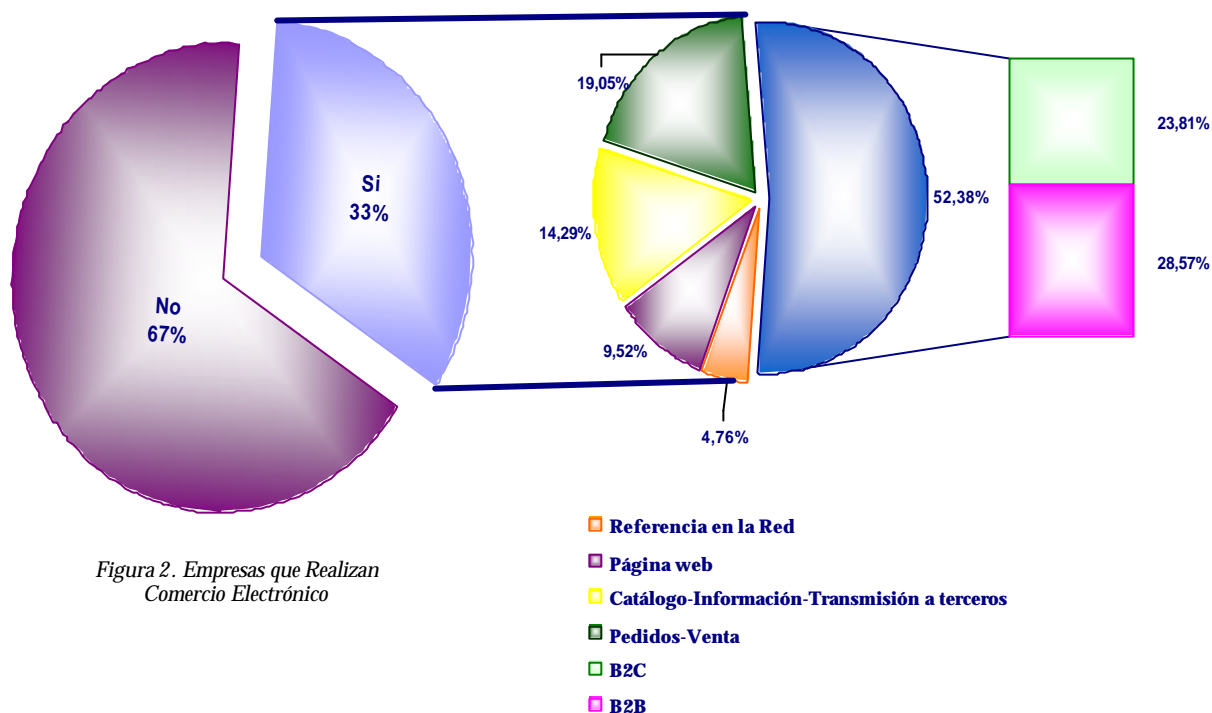


Figura 2. Empresas que Realizan Comercio Electrónico

Las empresas que indicaron no realizar nada respecto al tema consultado, afirmaron en un 56,27% no encontrarse preparadas para ello, situando su valoración por debajo del aprobado. Nuevamente surge la necesidad de romper con las barreras de la formación y del desconocimiento en nuevas tecnologías tal y como planteamos anteriormente. Considerando la desagregación por sectores que se ha realizado, podemos comprobar que pese a que las empresas del Sector Industrial no son las que más realizan actividades de comercio electrónico, sin embargo sí son las que otorgan un mayor grado de preparación sobre estos temas, seguido del Sector Servicios. Esta conclusión es lógica, ya que la especialización necesaria a nivel formativo y la estructura propia de los recursos en el Sector Industrial es mucho más elevada que la necesaria para el Sector Servicios, lo que hace que exista una disparidad patente entre ambos sectores.

Análisis de Regresión Logística. Probabilidad de que una Pyme Murciana realice Comercio Electrónico.

Cuando comenzamos a estudiar y analizar la situación de las empresas de la Región de Murcia, uno de los interrogantes principales que nos planteábamos era determinar no solamente su situación actual, sino conocer qué variables afectaban de forma determinante al hecho de que se realizara alguna actividad dentro de ellas que tuviera relación con el comercio electrónico. Aunque nuestro propósito inicial era determinar estas variables, no iba a ser posible conocer dicha relación hasta obtener los datos muestrales suficientes, y sólo si la respuesta de las empresas que realizaban algún tipo de política de comercio electrónico respecto a los que no, resultaba lo suficientemente representativa.

La pregunta planteada que nos indica si se realiza o no comercio electrónico en la empresa, tiene carácter dicotómico (denominada P6 a partir de ahora), por lo que al ser utilizada como variable dependiente en un análisis para estudiar qué variables influirían en su valor final, nos impedía que pudiéramos emplear técnicas de regresión habituales. Por este motivo decidimos utilizar un análisis de regresión logística, ya que este método nos permite pronosticar si se pertenece a un grupo o no en función de un grupo de variables independientes. No recurrimos al análisis discriminante porque éste no permite utilizar variables independientes categóricas, lo cual resultaba nuevamente un inconveniente ya que las variables relacionadas, o relacionables, con nuestro variable objeto de estudio en la encuesta aparecen evaluadas mediante escalas Likert, puesto que son estimadas de forma cualitativa por el individuo según las considere más o menos importantes. El análisis de regresión logística sí que nos permite utilizar variables independientes categóricas e incluir variables cuantitativas al mismo tiempo en el análisis, lo que la hace una técnica idónea a nuestros propósitos⁴.

En la estructura del cuestionario planteamos varios grupos de variables clasificadas según el tema a consultar que explicamos brevemente. Las preguntas 8 y 10, se refieren a la importancia de una lista de 12 objetivos y de 11 barreras respectivamente a la hora de que una empresa en la Región decida instaurar políticas de comercio electrónico. Por otro lado, las preguntas

⁴ Como herramienta de análisis se ha utilizado el *Paquete Estadístico SPSS 11.5 para Windows*.

9 y 11 instan a evaluar al encuestado, los factores más importantes como estímulos (7 variables) o barreras (10 variables) a la hora de que el consumidor se decida a realizar transacciones en la red. Consideramos que este punto de vista era también sumamente interesante incluirlo como variables que pudieran afectar a la hora de que una empresa se decida o no a efectuar Comercio Electrónico, ya que dependiendo de la visión que de éstos se tenga, probablemente se asuma o no el riesgo implícito de comenzar una forma de negocio diferente a la tradicional⁵.

Por otro lado, existen cuestiones de carácter general que describen el tipo de empresa sobre el que estamos trabajando y que consideramos importantes de introducir como variables independientes, tales como el Sector al que pertenece la empresa, la Categoría en la que se sitúa la misma y la Antigüedad.

El proceso para elaborar el análisis de regresión logística ha sido gradual. Se han realizado diversas regresiones, que de forma paulatina han ido incorporando variables predictoras a la regresión. Cuando decidimos emplear los datos de las variables independientes consideradas barreras generales en la Región, pudimos comprobar que ninguna de ellas se mantenía como variable en la ecuación final, por lo que finalmente se han desechado del análisis. Lo mismo ha ocurrido con la variable que señala la categoría de la empresa estudiada.

Las variables que nos proporcionan información acerca de los elementos que se consideran que pueden influir negativamente para que el consumidor compre, sí que han contribuido significativamente a la mejora en el modelo de regresión planteado. Sin embargo las variables propuestas acerca de los factores que inciden en el consumidor favorablemente para que éste compre por la red, han resultado ser no relevantes para nuestra regresión. Creemos que esta no inserción de las variables tales como comodidad, disminución del precio y del tiempo, personalización, anonimato, etc..., como covariables independientes de la regresión, puede estar ocasionado porque son

⁵ La nomenclatura utilizada para cada una de estas variables es el número de pregunta del cuestionario en el que se encuadran, más el número de ítem por orden de entrada en la lista; también se incluye una pequeña descripción para favorecer la comprensión de las mismas. Todas estas variables tienen un carácter categórico ya que todas han sido evaluadas mediante una escala Likert de 1 a 5.

percepciones que la empresa cree que tiene el consumidor y que influyen para incentivarle a realizar transacciones en la red, pero que realmente no son manejadas por la empresa, por lo que no van a tener un peso específico a la hora de incidir sobre si ésta realiza políticas de comercio electrónico o no. Sin embargo las barreras con las que se puede encontrar un consumidor a la hora de acceder a nuestro negocio en la red, o en realidad la percepción que de ellas tiene la empresa, sí que pueden resultar un motivo suficiente para que un negocio se instale en la web o no.

En la regresión logística se ha de decidir el método para seleccionar las variables que se incluirán o no en el modelo atendiendo a criterios teóricos o bien estadísticos. Hemos realizado todos los análisis utilizando diferentes métodos de introducción de variables y finalmente hemos optado por el método “hacia atrás” (*backward*) que excluye las variables basándose en la razón de verosimilitudes (*RV*). Este método recordamos que parte del *modelo saturado*, es decir, de aquel que contiene todas las variables que hemos seleccionado y que va excluyendo paso a paso las que no son significativas.

Todas las variables han sido consideradas como categóricas a excepción de la variable *Antigüedad* que es cuantitativa. El motivo es porque poniendo como ejemplo a la variable *Categoría de la Empresa*, el que la misma obtenga en un dato un valor de 1 y en otro un valor de 5 no significa que en este último caso quintuple su valor, sino que pertenecen a categorías diferentes de empresa, en el primer caso a una denominada grande, con más de 250 trabajadores y en el último a una empresa que sólo trabaja por la red, y que por lo tanto necesita de menos personal. Por este motivo, dichas variables necesitan ser categorizadas, es decir transformarlas a su vez en tantas variables *dummy* menos una, como categorías existan para dicha variable⁶.

El tipo de contraste utilizado para la categorización de las variables independientes ha sido el denominado *Indicador*, en la que cada categoría se compara con la categoría señalada de referencia mientras que esta última no se evalúa, ya que se le asignan ceros. Hemos utilizado como categoría de

⁶ Si la variable independiente solamente incorpora dos categorías, uno que indique ausencia de la característica y otra su inclusión, no es necesario hacer ninguna operación adicional. Pero éste no es nuestro caso, ya que todas las variables categóricas presentan cinco categorías con excepción de la variable *Sector Codificado* que cuenta con tres: 1. Servicios, 2. Industria y 3. Construcción.

referencia la última, debido a que expresamente queríamos introducir dentro del análisis el sector al que pertenece la empresa número uno, es decir la que nos arroja información acerca de las empresas pertenecientes al sector servicios, debido a la importancia y el peso que distintos estudios adjudican a este sector en la Región de Murcia.

Resumen del procesamiento de los casos

En la *Tabla 3* aparece la información sobre el número de casos válidos introducidos en el análisis y los excluidos. La

Casos no ponderados(a)		N	Porcentaje
Casos seleccionados	Incluidos en el análisis	434	96,4
	Casos perdidos	16	3,6
	Total	450	100,0
Casos no seleccionados		0	0
Total		450	100,0

a Si está activada la ponderación, consulte la tabla de clasificación para ver el número total de casos.

Tabla 3

codificación interna utilizada para separar la dicotomía de la variable dependiente ha sido un cero cuando

Historial de iteraciones(a,b,c)

Iteración		-2 log de la verosimilitud	Coefficientes
			Constante
Paso 0	1	553,021	-,664
	2	552,955	-,690
	3	552,955	-,690

a En el modelo se incluye una constante.

b -2 log de la verosimilitud inicial: 552,95

c La estimación ha finalizado en el número de iteración 3 porque las estimaciones de los parámetros han cambiado en menos de ,001.

Tabla 4

obtendremos será la Probabilidad de que la variable dependiente sea uno en función de una serie de variables independientes y una constante, es decir conseguiremos la probabilidad de que la empresa efectivamente realice alguna política de comercio electrónico, dadas una serie de circunstancias medidas por las variables independientes.

En la *Tabla 4* se presenta un resumen del modelo para el *Paso Cero*. Como ya sabemos, al no ser un modelo lineal se utiliza otro sistema para obtener los parámetros del mismo. El método utilizado de máxima verosimilitud es iterativo, por lo que se realiza por pasos que poco a poco incorporan mejoras en el ajuste respecto a un *modelo* denominado *nulo*. Es por esta razón que en el *Paso Cero* sólo se incluye la constante y no se introduce ninguna variable

No se realiza ninguna actividad de comercio electrónico en la empresa y un uno cuando Sí existe alguna política respecto a este tema. Es importante destacar que la categoría con el valor uno es imprescindible en la interpretación final del modelo de regresión, porque lo que finalmente

independiente. Los estadísticos de bondad de ajuste global para el modelo estimado solamente pueden ser tenidos en cuenta de forma orientativa debido a que la variable de objeto de estudio es una variable categórica, lo que puede llevar a que dichos coeficientes sean moderados dando la sensación de que dicho modelo no es adecuado cuando en realidad tal vez esto no sea así.

Pasando a los resultados obtenidos en los siguientes pasos, en primer lugar obtenemos una tabla denominada *historial de iteraciones* que en esta ocasión es muy extensa por lo que no la presentamos aquí, y en donde se van incorporando por el

Pruebas ómnibus sobre los coeficientes del modelo

		Chi-cuadrado	gl	Sig.
Paso 1	Paso	247,857	139	,000
	Bloque	247,857	139	,000
	Modelo	247,857	139	,000
Paso 2	Paso	-,689	4	,953
	Bloque	247,168	135	,000
	Modelo	247,168	138	,000
...

Paso 21	Bloque	-6502	4	,165
	Modelo	174,474	49	,000
	Paso	174,474	62	,000
Paso 22	Bloque	-6,157	4	,188
	Modelo	168,317	55	,000
	Paso	168,317	58	,000

Tabla 5

método de introducción de variables seleccionado las que se consideran significativas para el modelo y excluyendo las que no lo son. En total se han realizado 22 pasos hasta conseguir que el modelo no mejore en su ajuste por la inclusión de una nueva variable.

A continuación encontramos en la *Tabla 5* las pruebas denominadas *ómnibus* sobre los coeficientes del modelo o contrastes de ajuste global, pero en esta ocasión para cada uno de los pasos utilizados en la introducción de las variables. El valor de *Chi-cuadrado* permite contrastar que la mejora obtenida en el ajuste es nula, por lo que cada paso nos informa de cuál es la variación producida en el ajuste como consecuencia de la incorporación o eliminación de una variable nueva.

Es preciso tener en cuenta que el método *backward* con extracción por máxima verosimilitud utilizado, ya que por esta razón un valor de *Chi-cuadrado* negativo indica que ha disminuido este valor con respecto al paso anterior, lo cual en este tipo de metodología es lo correcto.

El modelo consigue realizar el mejor ajuste en 22 pasos, y se puede observar en la *Tabla 6* que presenta los estadísticos de ajuste global del modelo, que éste va mejorando progresivamente, disminuyendo a cada paso el valor de los mismos, conforme se eliminan variables no significativas del análisis. Hemos de fijarnos en esta referencia porque como ya se ha comentado antes, el valor de estos estadísticos de ajuste puede que

Estadísticos de ajuste global del modelo

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	305,098(a)	,435	,604
2	305,787(a)	,434	,603
...
20	371,979(b)	,341	,473
21	378,481(b)	,331	,460
22	384,638(b)	,321	,446

a La estimación ha finalizado en el número de iteración 20 porque se han alcanzado las iteraciones máximas. No se puede encontrar una solución definitiva.

b La estimación ha finalizado en el número de iteración 6 porque las estimaciones de los parámetros han cambiado en menos de ,001.

Tabla 6

nos lleve a cometer errores de apreciación, ya que pueden parecer elevados pese a que el ajuste sea bueno. El problema reside en la caracterización de la variable dependiente objeto de estudio, puesto que al ser dicotómica implica que la cuantificación de los estadísticos no sea la apropiada.

También se presentan los resultados del cálculo de la bondad del ajuste mediante la prueba de *Hosmer-Lemeshow*, estadístico muy utilizado en el análisis de

regresión logística.

Tabla de Contingencias - Prueba de Hosmer-Lemeshow

Suele ser muy útil en el caso de que existan numerosas variables independientes, lo que puede hacer que invalide la utilización de otros estadísticos de

	Paso	P6 Realizan o No CE = 0		P6 Realizan o No CE = 1		Total
		Observado	Esperado	Observado	Esperado	
	1	43	42,341	0	,659	43
	2	40	40,926	3	2,074	43
	3	38	39,246	5	3,754	43
	4	38	37,061	5	5,939	43
	5	37	33,886	6	9,114	43
	6	28	29,313	15	13,687	43
	7	26	25,190	17	17,810	43
	8	23	21,080	20	21,920	43
	9	8	14,350	35	28,650	43
	10	8	5,607	39	41,393	47

Tabla 7

ajuste más habituales. La *Tabla 7* de contingencias da origen al estadístico de *Chi-cuadrado de Hosmer-Lemeshow*, y en ella se presentan los deciles denominados de riesgo, en los que dicha prueba divide a la muestra a partir de

las probabilidades pronosticadas para cada uno de los pasos efectuados en el análisis.

Cada uno de estos deciles de riesgo indica el número de casos que pertenecen a cada categoría de la variable dependiente y el número de casos que para esa categoría pronostica el modelo obtenido en ese paso. Así conseguimos el número observado y el número pronosticado. Por comparación entre las observaciones y los valores esperados se calcula el estadístico *chi-cuadrado de Pearson*.

Prueba de Hosmer y Lemeshow

Paso	Chi-cuadrado	gl	Sig.
1	8,207	8	,414
2	5,695	8	,681
...
20	5,094	8	,748
21	5,203	8	,736
22	9,046	8	,338

Tabla 8

Mostramos en la *Tabla 8* los datos que presenta esta prueba para el Paso 22. Como se puede apreciar el estadístico toma un valor de 9,046 y tiene un nivel crítico de 0,338,

por lo que no se puede rechazar la hipótesis nula que plantea la igualdad de distribución de los datos en todos los deciles de riesgo en los que divide la muestra, por lo que podemos concluir que los

Tabla de clasificación-Matriz de Confusión

Observado			Pronosticado		
			P6 REALIZAN YA CE		Porcentaje correcto
			0	1	
Paso 1	P6 Realizan o No CE	0	259	30	89,6
		1	42	103	71,0
	Porcentaje global				83,4
Paso 2	P6 Realizan o No CE	0	260	29	90,0
		1	41	104	71,7
	Porcentaje global				83,9
...
Paso 21	P6 Realizan o No CE	0	263	26	91,0
		1	55	90	62,1
	Porcentaje global				81,3
Paso 22	P6 Realizan o No CE	0	261	28	90,3
		1	62	83	57,2
	Porcentaje global				79,3

Tabla 9

datos sí que se ajustan al modelo.

Al igual que antes con el resto de pruebas de bondad del ajuste, tenemos que indicar que el valor del estadístico puede ser muy elevado, ya que es sensible al tamaño de la muestra, pero al mismo tiempo la *prueba de Hosmer-Lemeshow* es más eficiente con muestras grandes, por lo que es un inconveniente difícil de salvar. Se genera así una nueva *matriz de confusión*

diferente a la obtenida en el *Paso Cero* que muestra el resultado de la clasificación para cada uno de los pasos⁷. Con este procedimiento, introduciendo todas las variables originales mencionadas, hemos obtenido el mejor porcentaje global llegando a explicar un 79,3% de todos los casos y el 90,3% de aquellas empresas que deciden no realizar ninguna política de comercio electrónico frente al 57,2% de las que eligen sí realizar alguna (*Tabla 9*).

Por último, se presenta un extracto de la tabla que recoge la información más relevante de este análisis, las variables incluidas en la ecuación con la estimación de los coeficientes, y el nivel crítico o de significación de las mismas (*Tabla 10*).

		Variables en la ecuación						I.C. 95,0% para EXP(B)	
		B	E.T.	Wald	gl	Sig.	Exp(B)	Inferior	Superior
Paso 22(a)	ANTIGÜED	0,024	0,010	6,553	1	0,010	1,025	1,006	1,044
	SECTORCO			10,212	2	0,006			
	SECTORCO(1)	1,757	0,568	9,548	1	0,002	5,793	1,901	17,651
	SECTORCO(2)	1,317	0,580	5,148	1	0,023	3,732	1,196	11,643
	P8.1_AHO			8,300	4	0,081			
	P8.1_AHO(1)	1,645	0,673	5,983	1	0,014	5,182	1,387	19,367
	P8.1_AHO(2)	-0,316	0,508	0,385	1	0,535	0,729	0,269	1,976
	P8.1_AHO(3)	0,111	0,447	0,062	1	0,804	1,118	0,465	2,685
	P8.1_AHO(4)	0,016	0,376	0,002	1	0,966	1,016	0,487	2,123
	P8.2_VEN			10,255	4	0,036			
	P8.2_VEN(1)	-1,580	0,895	3,114	1	0,078	0,206	0,036	1,191
	P8.2_VEN(2)	-1,285	0,579	4,921	1	0,027	0,277	0,089	0,861
	P8.2_VEN(3)	-0,475	0,435	1,196	1	0,274	0,622	0,265	1,457
	P8.2_VEN(4)	0,227	0,365	0,386	1	0,534	1,255	0,614	2,565
	P8.6_ADA			19,337	4	0,001			
	P8.6_ADA(1)	2,335	0,689	11,497	1	0,001	10,332	2,679	39,847
	P8.6_ADA(2)	2,158	0,603	12,816	1	0,000	8,654	2,655	28,205
	P8.6_ADA(3)	0,739	0,483	2,341	1	0,126	2,093	0,812	5,392
	P8.6_ADA(4)	1,127	0,458	6,054	1	0,014	3,087	1,258	7,576
	P8.7_VEN			33,551	4	0,000			
	P8.7_VEN(1)	-3,178	0,885	12,886	1	0,000	0,042	0,007	0,236
	P8.7_VEN(2)	-2,626	0,646	16,512	1	0,000	0,072	0,020	0,257
	P8.7_VEN(3)	-2,086	0,449	21,590	1	0,000	0,124	0,052	0,299
	P8.7_VEN(4)	-1,951	0,404	23,373	1	0,000	0,142	0,064	0,313

Tabla 10

Las variables que finalmente sí se encuentran en la ecuación son:

- a) De las variables consideradas de información general sobre la empresa:

⁷ Hemos utilizado el punto de corte habitual en este tipo de regresiones, que es del 0,5

- × **ANTIGÜED** ⇒ *Antigüedad* de la empresa, variable de tipo cuantitativo.
- × **SECTORCO** ⇒ *Sector Codificado*, sector al que pertenece la empresa⁸.

b) De las variables extraídas de la pregunta 8 que cuestionaba los objetivos más importantes para la empresa:

- × **P8.1_AHO** ⇒ Posibilidad de Ahorro en Costes.
- × **P8.2_VEN** ⇒ Venta
- × **P8.6_ADA** ⇒ Adaptación de la Demanda y Mejor Segmentación.
- × **P8.7_VEN** ⇒ Ventaja Competitiva ante los competidores.
- × **P8.10_BU** ⇒ Buscar Información.

c) De las variables extraídas de la pregunta 10 que valoraba los factores barrera de entrada para las empresas murcianas:

- × **P10.1_IN** ⇒ Inexperiencia en el ámbito de las Nuevas Tecnologías.
- × **P10.7_FA** ⇒ Falta de Desarrollo Tecnológico.
- × **P10.11_A** ⇒ Arraigo de la Cultura Empresarial reacia a los cambios.

d) De las variables extraídas de la pregunta 11 que valoraba de nuevo los factores barrera, pero en esta ocasión para que un consumidor se decida a realizar transacciones por la red:

- × **P11.1_DI** ⇒ Dificultad de acceso a Internet
- × **P11.2_LE** ⇒ Lentitud en el acceso a las direcciones buscadas.
- × **P11.3_CO** ⇒ Coste añadido de la Tecnología.
- × **P11.9_NO** ⇒ Incertidumbre sobre la Veracidad de la Información.
- × **P11.10_I** ⇒ Incertidumbre sobre cómo se Realizará el Proceso.

e) Más el término **Constante**.

⁸ Recordar que 1 era una empresa del sector Servicios, 2 representaba a una empresa del sector Industrial y 3 al sector Construcción.

Si el coeficiente de B es significativo, implica que la categoría a la que representa es bastante diferente a la categoría de referencia elegida. Por ejemplo, para la variable $P8.1_AHO$ que mide la variable independiente del objetivo de la empresa, de utilizarlo como nuevo sistema para incrementar las ventas, podemos observar que las variables $P8.2_VE(3)$ y $P8.2_VE(4)$ tienen unos niveles de significación muy altos, 0,274 y 0,534 respectivamente, por lo que podemos asegurar que la probabilidad de que la venta sea un objetivo muy importante no difiere de la probabilidad de la categoría de referencia.

Un signo negativo en los coeficientes de las variables implica que el incremento en cualquiera de ellas disminuye la probabilidad de que una empresa murciana realice comercio electrónico. Señalar además, que en este tipo de análisis, los coeficientes del modelo se expresan en las unidades originales de las variables independientes.

Además del nivel de significación hemos de tener en cuenta el valor del *odds ratio* calculado en la columna de $Exp(B)$. Los coeficientes positivos que sean significativos indicarán que la razón de ventajas será mayor a uno. Si comprobamos que el *odds ratio* es menor que uno, podemos indicar que la probabilidad de que sí realice comercio electrónico es menor en esa variable que en la categoría de referencia. Por ejemplo, la variable categórica $P11.1_DI(4)$, es decir aquella que mide en nuestro modelo a los sujetos que han dado una valoración de cuatro sobre cinco a la variable que estudiaba la *Dificultad en el Acceso a Internet* como barrera para el consumidor, obtiene un nivel crítico de 0,579, mientras que su razón de ventajas es de 0,764 lo que implica que la probabilidad de que la variable dependiente tome el valor uno, es decir de que la empresa que considera esta barrera con un valor de cuatro, es menor que la que indica que esta barrera obtiene una importancia de cinco, observación que parece lógica, ya que si la empresa considera que esta barrera es bastante importante, y que por lo tanto va a afectar a sus clientes potenciales de forma negativa, va a estar menos dispuesta a lanzarse a negociar por la red.

La significación de cada coeficiente se evalúa a partir del *estadístico de Wald* que también aparece en esta tabla, y que nos permite contrastar la hipótesis nula de que el coeficiente vale cero en la población. Éste se puede asimilar a

una χ^2 y tiene como inconveniente que es muy sensible al tamaño de los coeficientes, por lo que si su valor absoluto es muy alto resultará poco fiable.

Una vez obtenidos los coeficientes y las variables que sí se incluyen en el modelo, podemos construir una ecuación lineal (PR) que transformándola a la escala logística nos permitirá obtener las probabilidades pronosticadas.

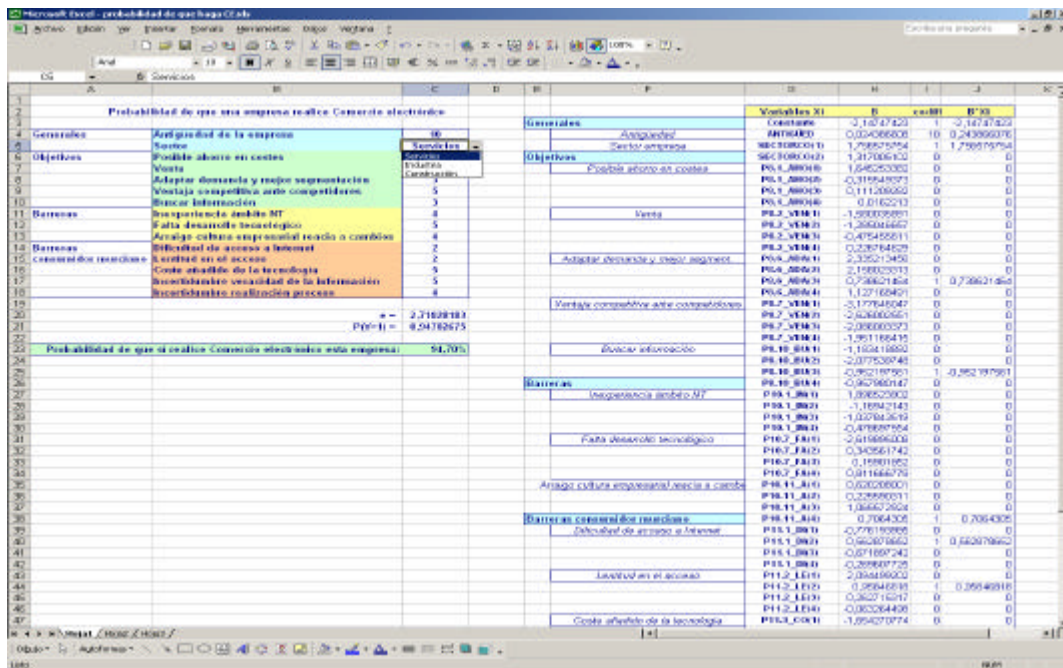


Figura 3 Cálculo de la Probabilidad

Para facilitar la estimación de dicha probabilidad, hemos planteado el modelo en una hoja de cálculo de Microsoft® Excel 2003 cuyo aspecto presentamos en la Figura 3, con el fin de que un empresario o la Administración disponga de una herramienta eficaz y sencilla de manejar para que introduciendo como datos iniciales los valores de las variables que se encuentran en el modelo, pueda realizar un pronóstico acerca de la probabilidad de que su empresa con esa percepción de las barreras y objetivos a alcanzar vaya a realizar comercio electrónico.

Por ejemplo, podemos conocer así que la probabilidad de que una empresa murciana del sector servicios, que tenga 10 años de antigüedad, realice comercio electrónico es del 77,43% (Figura 4) siempre que presente entre los objetivos a alcanzar mediante el comercio electrónico la venta, la ventaja competitiva que adquiere frente a los competidores y el posible ahorro en costes con una valoración máxima, mientras que le resulte más o menos

indiferente el objetivo de búsqueda de información y la adaptación de la demanda, que valore con un 5 en su empresa la barrera de falta de desarrollo tecnológico, y con un 4 el arraigo de la cultura empresarial reacia en cambios y la inexperiencia en el ámbito de las nuevas tecnologías, mientras que considere muy importantes las barreras que tienen los consumidores por el coste añadido de la tecnología y la incertidumbre que les genera la realización del proceso, creyendo que es un poco menor la inseguridad en la veracidad de la información y valorando con un dos la dificultad de acceso y la lentitud de la red.

Probabilidad de que una empresa realice Comercio Electrónico

Generales	Antigüedad de la empresa Sector	10 Servicios
Objetivos	Posible ahorro en costes	5
	Venta	5
	Adaptar demanda y mejor segmentación	3
	Ventaja competitiva ante competidores	5
	Buscar información	3
Barreras empresa	Inexperiencia ámbito NT	4
	Falta desarrollo tecnológico	5
	Arraigo cultura empresarial reacia a cambios	4
Barreras consumidor murciano	Dificultad de acceso a Internet	2
	Lentitud en el acceso	2
	Coste añadido de la tecnología	5
	Incertidumbre veracidad de la información	4
	Incertidumbre realización proceso	5
Probabilidad de que sí realice Comercio electrónico esta empresa:		77,43%

Figura 4

Sin embargo una empresa de la misma antigüedad, pero que pertenezca al sector industrial y que mantiene las mismas consideraciones y los mismos planteamientos sobre los objetivos y barreras introducidos por el modelo obtiene una probabilidad de realizar comercio electrónico de un 68,86%, mientras que si pertenece al sector de la Construcción esta probabilidad se reduce aún más situándose en el 37,20%.

Conclusiones

A modo de conclusión podemos afirmar que mediante este trabajo de investigación sobre una muestra de empresas importante de la región de Murcia, se ha determinado en qué estado se encuentra el comercio electrónico en las mismas y además se ha elaborado un modelo que nos indica que de forma general, las empresas del sector servicios son las que efectivamente presentan la probabilidad más alta de realizar actividades de comercio

electrónico, seguidas de las del sector industrial y por último el sector de la construcción, al tiempo que obtienen probabilidades más elevadas las empresas de mayor antigüedad, y las que menos valoraban la falta de veracidad del proceso o de la información, y en general las barreras a las que se enfrenta el consumidor, frente a las que consideraban con un alto grado de valoración las barreras propias que tienen que salvar. Indicar que bastantes de las empresas que afirmaron no realizar políticas de comercio electrónico en la muestra, si calculamos con sus datos la probabilidad de que sí pudieran realizar alguna actividad en torno a este tema, los resultados no son excesivamente bajos siempre y cuando ponderen como un objetivo importante a alcanzar, la Venta y sostener una Ventaja Competitiva frente a los demás, además de los ítems anteriormente comentados.

Bibliografía

- [1] CARM. "Economía de la Región de Murcia". www.carm.es/carm/region/economia/economia.jsp
- [2] *Diario electrónico "Economía Regional"*. "El modelo económico de la Región de Murcia" 30 de Octubre de 2001.
- [3] Directorio de empresas. Cámaras de Comercio. <http://directorio.cameras.org>
- [4] *Gondar Nores, J.E.* Data Mining Intitute SL "SPSS y Modelado Estadístico" www.estadistico.com/arts.html?20001113
- [5] *Greene, W. H.* "Análisis Econométrico". Ed. Prentice Hall. Madrid, 1998.
- [6] *Gujarati, D.* "Econometría". 2ª Edición. McGraw-Hill. 1981.
- [7] *Hosmer, D.W y Lemeshow, S.* "Applied Logistic Regresión". Ed. John Wiley. Nueva York, 1989.
- [8] *Lizasoain, L; Joaristi. L.* *Gestión y análisis de datos con SPSS versión 11. Universidad del País Vasco. Ed. Thomson-Paraninfo. Madrid, 2003.*
- [9] *Múria Albiol, J. Gil Saura, R.* "Preparación, Tabulación y Análisis de Encuestas para Directivos." *ESIC Editorial, 1998.*
- [10] *Muñoz Reyes, A. M, Jiménez Gomero, M.D, Muñoz García, J., Pino Mejías, R.* "e-Encuestas Probabilísticas II. Los métodos de Muestreo Probabilística" *Estadística Española, Vol. 44, núm 151, 2002*
- [11] *Novell, A. J.* "Análisis de Regresión Logística". Cuadernos Metodológicos. CIS. 1995.
- [12] *Pardo Merino, A y Ruiz Díaz, M.A.* "SPSS 11. Guía para el análisis de datos". Ed. McGraw-Hill, 2002.
- [13] *Pérez, C.* "Técnicas Estadísticas con SPSS". Ed. Prentice Hall, Madrid, 2001.
- [14] *Santos Peña, J., Muñoz Alamillos, A. Juez Martel, P. y Guzmán Justicia, L.* "Diseño y tratamiento estadístico de encuestas para estudios de mercado". Ed. Centro de Estudios Ramón Areces, SA. 1999.
- [15] *Takeshi, A.* "Modelos de respuesta cualitativa: un examen" Cuadernos Económicos del ICE. Nº 39. 1988/2.