



---

**Universidad de Valladolid**  
Facultad de Ciencias

## **TRABAJO FIN DE GRADO**

Grado en ESTADÍSTICA

# **Análisis de oportunidad de compra para los socios de un Club de Fidelidad**

Convocatoria Septiembre de 2014

*Autor:* Verónica Pozo Martín

*Tutores:* Eusebio Arenal Gutiérrez  
Aníbal Bregón Bregón

# ÍNDICE

1. INTRODUCCIÓN .....	1
2. RESUMEN .....	2
3. ABSTRACT .....	3
4. DATOS	
4.1.    Recopilación de los datos .....	6
4.2.    Tratamiento y depuración de los datos .....	8
4.3.    Estructura de la base de datos .....	9
4.4.    Estructura de las tablas y tipos de datos .....	10
5. RESULTADOS	
5.1.    Determinar muestra representativa .....	14
5.2.    Perfil socios de la base de datos .....	15
5.3.    Comportamiento de compra .....	23
5.4.    Segmentación .....	29
5.5.    Análisis de resultados .....	32
6. CONCLUSIONES .....	43
7. BIBLIOGRAFÍA .....	44
8. LISTA DE FIGURAS Y TABLAS .....	45
9. ANEXOS .....	47

# 1. INTRODUCCIÓN

La empresa a la que hace referencia este trabajo, es una marca de moda a cuyos clientes se les da la posibilidad de pertenecer a un club de fidelidad. La tarjeta del Club ofrece a sus socios una serie de ventajas en cuanto a forma de pago, fidelización y servicios, así como un trato favorable a cerca de la información sobre promociones y descuentos. La misión del club de fidelidad es crear mediante la Tarjeta Club relaciones comerciales y emocionales duraderas y beneficiosas para ambas partes.

Actualmente, las promociones son principalmente genéricas y dirigidas sobre el total de clientes que se tiene, donde el incentivo es prácticamente igual para todos. Pero los resultados de los test promocionales realizados en este ejercicio y anteriores detectan que la mayoría de los clientes quiere que se premie su lealtad, de manera que prefieren los programas en los que el grado de fidelidad influye realmente para conseguir mejores ofertas y servicios. También desean ser tratados de manera diferente y recibir un servicio mejor del que reciben los compradores que no son miembros del Club.

Entre los objetivos del Club están construir un concepto de fidelización con valores añadidos que nos diferencie de los existentes, para no ser uno más, que sea afín a nuestro cliente, a su estilo de vida, etc., y que refuerce el posicionamiento de la marca. Se pretende mantener una relación dinámica, joven, viva, que lleve a marca a ser parte de la vida del cliente...que la compra le proporcione “más satisfacciones que la mera compra de ropa”.

Para ello, se propone crear una campaña personalizada al socio en función de su compra, con el fin de afianzar su confianza en la marca, y, por otro lado, aumentar el ticket medio y las operaciones realizadas.

Para decidir la campaña, se va a realizar un estudio sociodemográfico y de comportamiento de compra previos. A continuación se analizarán los resultados de dicha campaña para obtener conclusiones que se tendrán en cuenta para futuras acciones por parte del Club a sus socios.

## 2. RESUMEN

En este trabajo se analizarán los datos que se obtienen de los socios con respecto a la información sociodemográfica, así como su comportamiento de compra, para mejorar las campañas. Interesa realizar segmentaciones y obtener información más adaptada a cada uno de los socios para poder personalizar las comunicaciones que se le envían, aumentar el número de las operaciones que realizan con respecto a temporadas anteriores y obtener un gasto medio por socio más elevado.

A lo largo de este trabajo se desarrollará la forma de obtención de la información, la estructura de la misma para un correcto tratamiento de los datos y la obtención de resultados fiables.

A continuación se realizará un análisis sociodemográfico de los socios, así como su comportamiento de compra, con el fin de seleccionar un conjunto que represente a todo el conjunto y que se utilizará como test para futuras acciones. A partir de estos resultados, se realizará una campaña a nivel de producto, que posteriormente se analizará para concluir si es favorable extenderlo al resto de la población.

La Dirección de la empresa, ha establecido una serie de políticas para realizar la segmentación. Fueron decididas por la experiencia comercial de la que se dispone. Se comprobará a lo largo del trabajo si estadísticamente estas políticas son las adecuadas.

La primera de las políticas será tomar las regiones de Asturias y Murcia, para realizar el test. Se estudiará si las dos zonas elegidas representan al conjunto de España mediante un análisis descriptivo de los socios, y un contraste de hipótesis sobre las principales variables sociodemográficas y de compra.

Por otro parte se dispone de diferentes familias de producto con las que se va a incentivar a los socios. Se realizará un modelo de árbol de decisión para extraer las familias de producto que se deben considerar en la segmentación.

A continuación, se preparará la campaña que se enviará al socio. Se realizará la segmentación en base al comportamiento de compra en los grupos de familia establecidos. Dentro de la segmentación, se crearán grupos de control y de espejo, aplicando un muestreo aleatorio simple, necesarios para ver la rentabilidad de la campaña.

Por último, se realizará un análisis de la campaña, aplicando un test de significatividad a través de un contraste de hipótesis, y se extraerán conclusiones de los resultados obtenidos.

Se utilizará MicroStrategy como herramienta para la elaboración de informes, *Teradata Relationship Manager*, es el gestor con el que se elaborarán las campañas, y a través de *SPSS Modeler* se creará el modelo de árbol de decisión.

### 3. ABSTRACT

In this paper the data obtained from partners regarding sociodemographic information, as well as their buying behavior will be analyzed to improve the campaigns. Segmentations and interested in making information more accessible to each of the partners to personalize communications sent to it, increasing the number of operations performed over previous seasons and get a higher average expenditure per partner.

Throughout this paper, the method of production information is developed, the same structure for a correct treatment of the data and to obtain reliable results.

Below is a demographic analysis of the partners and their buying behavior in order to select a set representing the whole and to be used as test for future actions will be performed. From these results, a campaign-level product, which is later analyzed to conclude whether it is favorable to extend to the rest of the population is performed.

The management of the company has established a series of policies for segmentation. Were determined by commercial experience that are available. It will check over work if statistically these policies are appropriate.

The first policy is to take the regions of Asturias and Murcia, to perform the test. It will examine whether the two selected areas represent all of Spain by a descriptive analysis of the partners, and a hypothesis test on the main sociodemographic variables and purchase.

On the other hand you have different product families which will encourage partners. A model of decision tree will be made to remove the product families to consider segmentation.

Then the campaign to be sent to the partner will be prepared. Segmentation will be made based on the purchase behavior of family groups established. In the segmentation, and control groups mirror will be created by applying a simple random sampling, necessary for the performance of the campaign.

Finally, an analysis of the campaign will be implemented by applying a test of significance through a hypothesis test, the results and conclusions obtained are extracted.

MicroStrategy will be used as a tool for reporting, Teradata Relationship Manager is the manager with campaigns, and through SPSS Modeler model is created decision tree is developed.

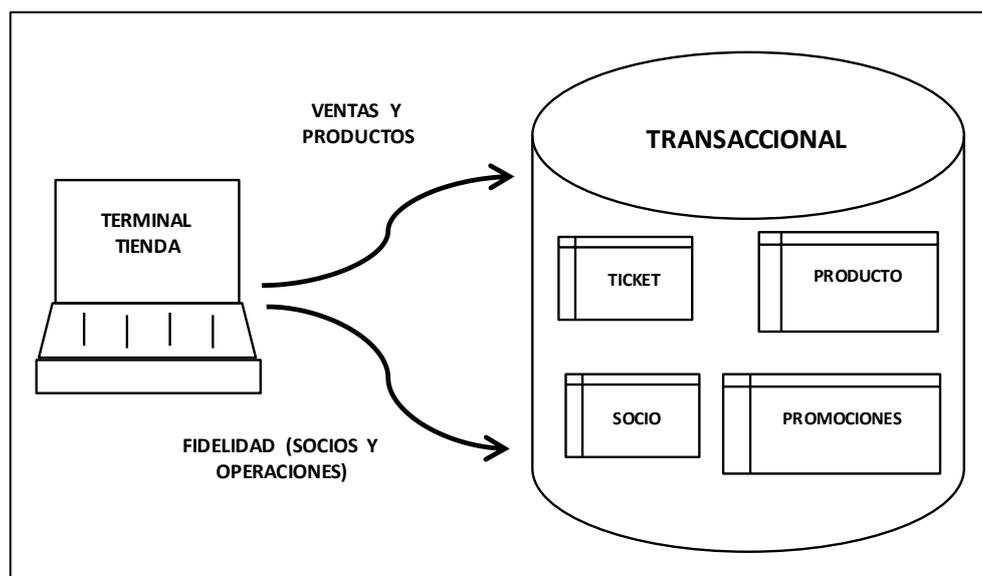
## 4. DATOS

En este capítulo se describirá cómo se alimenta la base de datos con la que se va a trabajar, de dónde proceden los datos que la forman, cómo son tratados y depurados, y la estructura de tablas donde se recogen.

### 4.1. Recopilación de los datos

Los datos que alimentan la base de datos se recogen principalmente en las tiendas; provienen de dos fuentes diferentes: por un lado está toda la información recogida a través de las operaciones realizadas por los clientes en las tiendas, y por otro lado, los datos recogidos mediante un cuestionario que rellenan los clientes que quieren hacerse socios del club de fidelidad.

Esquema de la recopilación de los datos de partida



(Figura 1)

Los datos obtenidos de las transacciones realizadas en tienda, se recogen a través del terminal de la misma y cada noche se transmiten a la base de datos de la central, que se denomina "Transaccional". El ticket de compra proporciona los datos que se van guardar.

A partir del ticket de compra se obtienen los siguientes datos que identifican la operación: la tienda donde se ha realizado la compra o devolución, la caja dentro de la tienda, fecha y hora de la operación, así como la información de producto, código, descripción, cantidad y precio del artículo adquirido.

Imagen de un ticket de compra



(Figura 2)

Estos datos de compra serán grabados en una tabla global dentro del transaccional que se denominada Ticket, sin pasar por ningún tipo de depuración ni manipulación. De esta manera, se dispone de datos de productos vendidos y ventas realizadas por todos los clientes que operan.

Los datos de fidelidad son los datos del socio recogidos mediante el cuestionario de solicitud con el que entran a formar parte del Club de Fidelidad; se transmiten de manera inmediata a la base de datos y son guardados en la tabla global Socios. Su tratamiento requiere el consentimiento del socio con el fin de cumplir con lo establecido en la Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal.

Imagen de un cuestionario de solicitud



(Figura 3)

## Imagen de una Tarjeta de Fidelidad



(Figura 4)

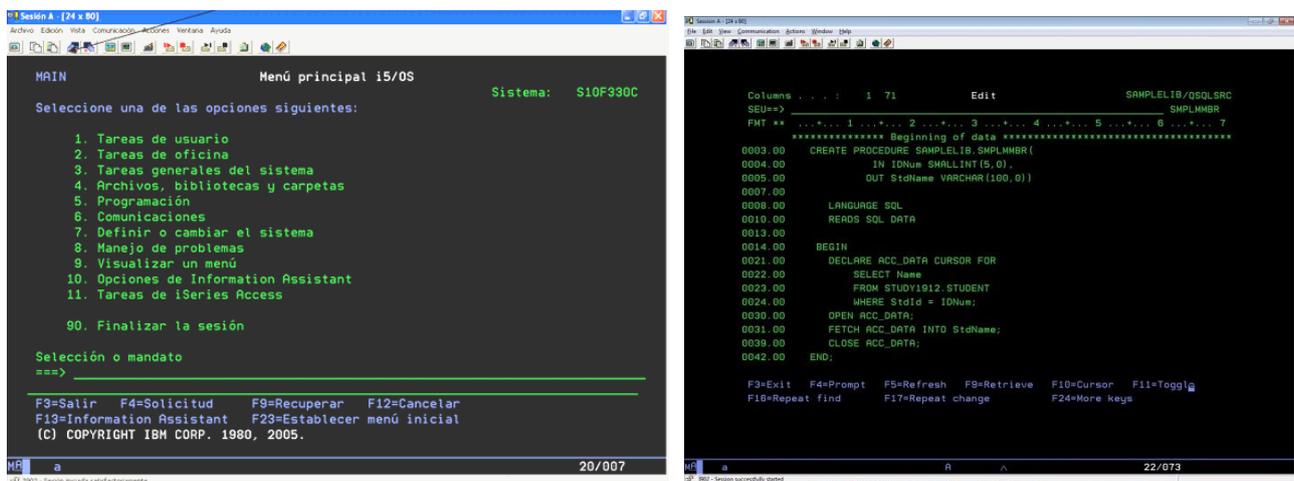
En el Transaccional se almacena, además, toda la información referente a producto, en la tabla Producto, y la de las promociones que se realizan a los socios o en los escaparates, recopilada en la tabla denominada Promociones y es alimentada por el departamento de ventas.

En resumen, lo que se tiene hasta este momento son los datos de partida almacenados en cuatro tablas diferentes del Transaccional, a partir de las cuales se va a formar la base de datos manipulada y depurada, con la que se va a trabajar.

## 4.2. Tratamiento y depuración de los datos

La base de datos del Transaccional, está soportada sobre un sistema AS/400 (equipo de IBM), que es una base de datos con una interfaz controlada mediante menús y comandos. Se accede mediante lenguaje de programación y permite, entre otros, el manejo de objetos grandes, la definición de datos y funciones, soporte multimedia.

### Interfaz del sistema AS/400



(Figura 5)

Los datos del Transaccional son manipulados a través de diversos procesos con el fin de prepararlos para ser analizados. Estos procesos se conocen como procesos ETL y constan de tres fases: extracción (**E**xtract), transformación (**T**ransform) y carga (**L**oad).

El proceso de extracción consiste en la recuperación de la información del Transaccional. Posteriormente, mediante el proceso de transformación, los datos son analizados a través de un chequeo para verificar que son correctos. Este chequeo se lleva a cabo mediante diferentes procesos de depuración:

- Se comprueba que la fecha y hora del ticket son válidos, de manera que si por algún motivo la caja pierde la conexión en el momento de realizar la operación, se le asigna una determinada.
- Se comprueba que la tienda, también llamada sucursal, existe, en caso contrario, se le asigna una sucursal ficticia.
- Se verifica que el producto que se ha marcado es correcto; es posible que se haya tecleado a mano y no exista, en cuyo caso se le asignará un número de referencia o patrón genérico.
- Sobre los campos de e-mail y teléfono se aplican distintas reglas que definen si esa dirección de correo electrónico o ese número de teléfono móvil son correctos. En el caso del e-mail se especifica que la longitud sea 2+1@+2+1.+2 (los números son número mínimo de caracteres), y para el teléfono se exige un ancho fijo de 9 caracteres, sin puntos, sin blancos y que comience por 6 o por 7, en el caso de España.
- El campo DNI se valida verificando que su largo es de 8 caracteres, que contiene una letra al final, y además, se diferencia a los DNI denominados de extranjeros, por su diferente nomenclatura.

Por último, el proceso de carga, a través del cual los datos son cargados en el sistema de destino. Llamamos sistema de destino al *Datawarehouse* de la empresa, es decir, la base de datos con una estructura homogénea y fiable, diseñada para favorecer el análisis.

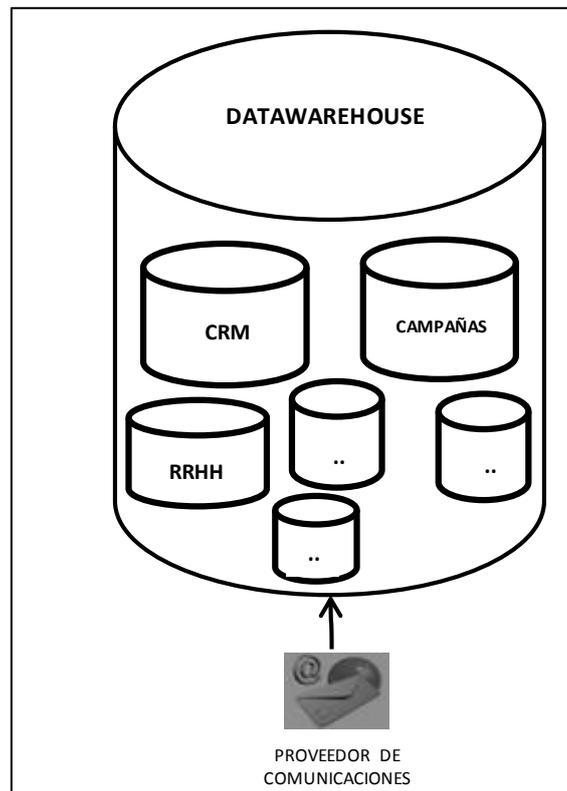
### **4.3. Estructura de la base de datos**

Se cuenta con una gran base de datos de clientes potenciales y actuales, lo que es fundamental para crear un entorno adecuado desde el cual se gestiona y administra toda la información de la gestión de ventas y de los clientes de la empresa.

Con el propósito de ayudar a que un área específica dentro del negocio pueda tomar mejores decisiones, se crean subconjuntos de datos, conocidos también como *data Marks*, de forma que los usuarios puedan tratar la información de manera individualizada y manipular solamente la información que concierne a su área.

En concreto, el subconjunto de datos que se utilizará se denomina CRM (*customer relationship management*), y se caracteriza principalmente por la posibilidad de validar directamente la información, realizar consultas sencillas y rápidas o la facilidad para la historización de los datos.

## Estructura del Datawarehouse



(Figura 6)

El CRM está orientado a la fidelización de los clientes, lo que permite disponer de información actualizada sobre ellos. Ayuda, además, a conocer los diferentes puntos de contacto con los cuales el cliente interactúa con los usuarios. Objetivos fundamentales son atraer nuevos clientes, retener a los actuales, recuperar antiguos y reducir costes de marketing.

El Datawarehouse es alimentado además, a partir de las comunicaciones que se envían a los socios; los proveedores de comunicaciones reportan regularmente el estado de los envíos realizados, de manera que se mantienen actualizados los *flags* o parámetros de control de la dirección, e-mail y teléfono, que indican si la información que se tiene es válida o inválida, esto es, si los socios son comunicables o no por cada canal.

En la base de datos se almacena también información de todos los departamentos de la empresa, recursos humanos, formación, contabilidad, producto, compras, etc., información que se va recopilando e integrando diariamente.

Finalmente, para que la base de datos cumpla con los requerimientos adecuados para obtener la máxima calidad de los datos, se aplican una serie de reglas a las tablas con el fin de evitar la redundancia de datos, disminuir los problemas de actualización de las tablas y proteger la integridad de los datos.

#### 4.4. Estructura de las tablas y tipos de datos

La base de datos utiliza tablas para guardar, organizar y presentar la información. Los atributos que se recogen en dichas tablas se caracterizan por el grupo de “tipo de datos” al que pertenecen, esto es, el conjunto de valores que puede tomar de manera que si se intenta dar un valor fuera de ese conjunto, se producirá un error.

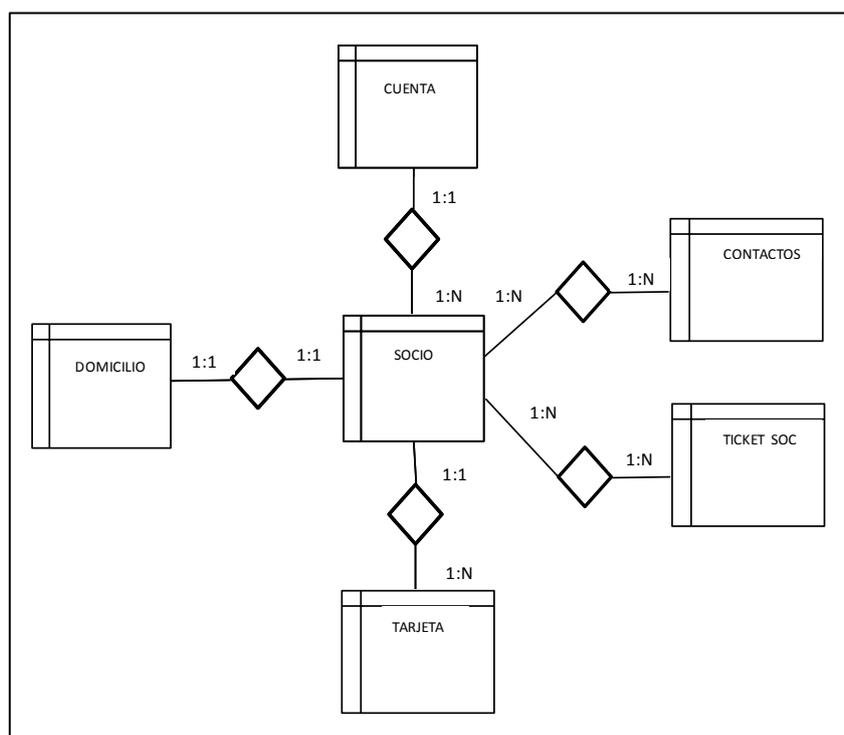
La clasificación de los tipos de datos de las variables que se van a utilizar en nuestro análisis se resumen en esta tabla:

NOMBRE DEL TIPO	DESCRIPCIÓN
CHAR	Texto de tamaño fijo
VARCHAR	Texto de tamaño variable
INTEGER	Entero de tamaño fijo
SMALLINT	Entero de 5 cifras
INTEGER	Entero de 10 cifras
DECIMAL	Real que puede trabajar con más decimales de los fijados
DATE	Almacena valores del tipo día, mes y año
TIME	Almacena valores del tipo hora, minuto y segundo
BOOLEAN	Tipo lógico (verdadero/falso)

(Figura 7)

En cuanto a la estructura de tablas del CRM, se basa en un modelo de entidad-relación que permite representar las entidades relevantes así como sus interrelaciones y propiedades.

#### Modelo Entidad-Relación



(Figura 8)

Mostramos las principales tablas que se van a utilizar para elaborar la campaña y consultas, y mediante qué campo o campos están unidas entre sí:



(Figura 9)

NOTA: identificamos con xx la longitud de caracteres que ha de tomar el campo en cada caso.

El campo común que une las tablas es el campo CTA\_ID, que es el campo cuenta, e identifica a cada uno de los socios pertenecientes al Club de Fidelidad.

En cuanto a la información que recogen las tablas, en la tabla SOCIO están recogidos los datos personales y de domicilio del socio, junto con toda la información referente a su registro como fecha de alta o baja y permiso de comunicaciones. Otro tipo de información relacionada con el socio, reside en otras tablas: en la tabla TARJETA está recogido si se encuentra de alta, de baja o en un estado intermedio; en la tabla DOMICILIO está toda la información referente a su lugar de residencia; en la tabla CONTACTOS podemos encontrar todas las comunicaciones y promociones que se le han comunicado, así como el canal de comunicación por el que se le ha enviado; y en la tabla TICKET\_SOC se localizan las compras realizadas por el socio así como información identificativa de la operación, el medio de pago con el que la realizó, el importe, la sucursal o si se le aplicó alguna promoción.

Por otro lado, en la base de datos se tienen una serie de tablas asociadas a las tablas del socio, en las cuales se recogen los diferentes estados o valores que pueden tomar los atributos; se muestran las que se van utilizar en este trabajo:

SEGMENTO	
SEGMENTO_ID (SMALLINT)	
SEGMENTO_DESC (VARCHAR xxx)	
SEGMENTO_ID	SEGMENTO_DESC
-1	Desconocido
1	Activo 0-6
2	Activo 7-12
3	Activo 2 años
4	Activo 3 años
5	Activo 4 años
6	Activo 5 años
7	Activo + de 5 años
8	Abandonista Reciente 2 años
9	Abandonista Reciente 3 años
10	Abandonista Reciente 4 años
11	Abandonista Reciente 5 años
12	Abandonista Reciente + de 5 años
13	Abandonista 24, 3 años
14	Abandonista 24, 4 años
15	Abandonista 24, 5 años
16	Abandonista 24, + de 5 años
17	Abandonista 4 años
18	Abandonista 5 años
19	Abandonista + de 5 años
20	Inactivo 0-6
21	Inactivo 7-12
22	Inactivo 2 años
23	Inactivo 3 años
24	Inactivo 4 años
25	Inactivo 5 años
26	Inactivo + de 5 años

TARJ_ESTADO	
TARJ_ESTADO_ID (CHAR x)	
TARJ_ESTADO_DESC (VARCHAR xx)	
TARJ_ESTADO_ID	TARJ_ESTADO_DESC
1	ALTA
2	BAJA
3	ESPERA
4	OTROS

SITUACION_SOCIO	
SITUACION_SOCIO_ID_ID (CHAR x)	
SITUACION_SOCIO_DESC (VARCHAR xx)	
TARJ_ESTADO_ID	TARJ_ESTADO_DESC
1	TITULAR
2	BENEFICIARIO
3	TARJETA FICTICIA
4	PROSPECT

SEXO	
SEXO_ID (CHAR x)	
SEXO_DESC (VARCHAR xxx)	
SEXO_ID	SEXO_DESC
M	MUJER
H	HOMBRE
D	DESCONOCIDO

MEDIOS_PAGO	
MEDIO_PAGO_ID (SMALLINT)	
MEDIO_PAGO_DESC (VARCHAR xxx)	
MEDIO_PAGO_ID	MEDIO_PAGO_DESC
-1	DESCONOCIDO
1	EFFECTIVO
2	PAGO CON SEÑAL
3	TARJETA CLUB
4	TARJETA VISA
5	EMPLEADOS
6	TALONES

ARREGLOS	
ARREGLOS_ID (SMALLINT)	
ARREGLOS_DESC (VARCHAR xxx)	
ARREGLOS_ID	ARREGLOS_DESC
-1	DESCONOCIDO
1	PANTALONES
2	PIEL
3	AMERICANAS
4	OTROS

AUTORIZA_MAIL	
AUTORIZA_MAIL_ID (SMALLINT)	
AUTORIZA_MAIL_DESC (VARCHAR 100)	
AUTORIZA_MAIL_ID	AUTORIZA_MAIL_DESC
0	Autoriza todos
1	No autoriza Grupo
2	No autoriza Terceros
3	No Grupo ni Terceros
4	No autoriza ninguno

AUTORIZA_EMAIL	
AUTORIZA_EMAIL_ID (SMALLINT)	
AUTORIZA_EMAIL_DESC (VARCHAR 100)	
AUTORIZA_EMAIL_ID	AUTORIZA_EMAIL_DESC
0	Autoriza todos
1	No autoriza Grupo
2	No autoriza Terceros
3	No Grupo ni Terceros
4	No autoriza ninguno

AUTORIZA_SMS	
AUTORIZA_SMS_ID (SMALLINT)	
AUTORIZA_SMS_DESC (VARCHAR 100)	
AUTORIZA_SMS_ID	AUTORIZA_SMS_DESC
0	Autoriza todos
1	No autoriza Grupo
2	No autoriza Terceros
3	No Grupo ni Terceros
4	No autoriza ninguno

(Figura 10)

NOTAS:

1. La tabla SEGMENTO define la variable que identifica el valor del socio; se calcula a partir de las compras realizadas por el socio en el último año.
2. El estado de la tarjeta describe si un socio está dado de alta, dado de baja o se encuentra en un estado intermedio, por ejemplo, pendiente de firma, en espera de recibir un nueva por pérdida o robo.
3. La situación del socio identifica al tipo de socio, esto es, si es titular de la tarjeta, beneficiario de este, tarjeta ficticia asignada o prospect, que es un socio potencial.
4. Las tablas de Autoriza son similares, solo se identifican por el permiso por parte del socio para ser comunicado por mail o correo postal, por e-mail o por sms.

## 5. RESULTADOS

En primer lugar, se va a determinar el tamaño de muestra apropiado para que sea fiable y representativo de la población que se tiene, junto con un perfil general de los socios de las zonas de test y del conjunto de España, con el fin de deducir si la muestra de Asturias y Murcia como test es válida.

A continuación se creará la segmentación de la campaña que se quiere comunicar y, por último, se realizará un análisis de los resultados obtenidos.

### 5.1. Determinación de una muestra representativa

Se va a determinar el tamaño de la muestra para decidir si las zonas de test elegidas pueden utilizarse.

Para que una muestra sea fiable, es necesario que su tamaño sea obtenido mediante procesos matemáticos que eliminen la incidencia del error.

Para calcular el tamaño de la muestra suele utilizarse la siguiente fórmula: (Ver Anexo 1)

$$n = \frac{N\sigma^2Z^2}{(N-1)e^2 + \sigma^2Z^2}$$

Donde:

n = el tamaño de la muestra.

N = tamaño de la población.

$\sigma$  = Desviación estándar de la población que, generalmente cuando no se tiene su valor, suele utilizarse un valor constante de 0,5.

Z = Valor obtenido mediante niveles de confianza. Es un valor constante que, si no se tiene su valor, se lo toma en relación al 95% de confianza equivale a 1,96 (como más usual) o en relación al 99% de confianza equivale 2,58, valor que queda a criterio del investigador.

e = Límite aceptable de error muestral que, generalmente cuando no se tiene su valor, suele utilizarse un valor que varía entre el 1% (0,01) y 9% (0,09), valor que queda a criterio del encuestador.

Por tanto, en este caso se tiene:

$$N = 3.363.522$$

$$\sigma = 0,5$$

$$Z = 1,96$$

$$e = 0.05$$

$$\Rightarrow n = 9.576$$

Por tanto, la muestra que se ha elegido de las regiones Asturias y Murcia, que supera este valor, con lo que en principio sí sería muestra suficiente para realizar el análisis.

## 5.2. Perfil socios de la base de datos

En este apartado se realizará un estudio de las variables principales que se tienen de los socios. Se estudiarán por separado las zonas de Asturias y Murcia, para verificar si resultan representativas para el conjunto de España y responder así a la pregunta de si la elección de estas zonas para realizar la campaña de test es acertada.

Por un lado, se verá cómo se comportan las variables cualitativas sexo, tramo edad y segmento, que provienen de la tabla Socio. Y por otro lado, se verá también el comportamiento de algunas métricas de compra calculadas a partir de los datos que se tienen en la tabla Ticket: Venta Neta (VN), Ticket Medio, Media de Compra, Frecuencia de Compra, Artículos por Operación y número de Operaciones, que se definirán más adelante.

### ○ Variables sociodemográficas:

Para las variables sexo, tramo\_edad y segmento, se realizará un histograma de frecuencias para observar su comportamiento, que irá acompañado de un contraste de bondad de ajuste de Chi cuadrado en el que se tomará como hipótesis nula que la variable estudiada es independiente de la zona geográfica de la que se trate, es decir, se comporta igual en las regiones de Asturias y Murcia que en el total de España.

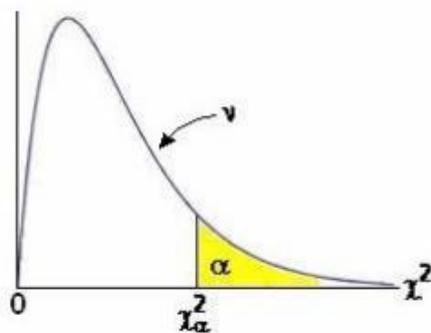
Para realizar el contraste se tiene que, dado un nivel de significancia  $\alpha$ , se define un valor crítico  $\chi^2_{\alpha}$  para el rechazo de la hipótesis propuesta:

$$H_0: f(x) = f_0(x)$$

Si las frecuencias observadas, correspondientes a Asturias y Murcia, no difieren significativamente de las frecuencias esperadas, correspondientes a toda España, entonces el valor de estadístico de prueba  $\chi^2$  será cercano a cero, pero si estas diferencias son significativas, entonces el valor del estadístico  $\chi^2$  estará en la región de rechazo de  $H_0$ :

$$\text{rechazo } H_0 \Leftrightarrow \chi^2 > \chi^2_{\alpha}$$

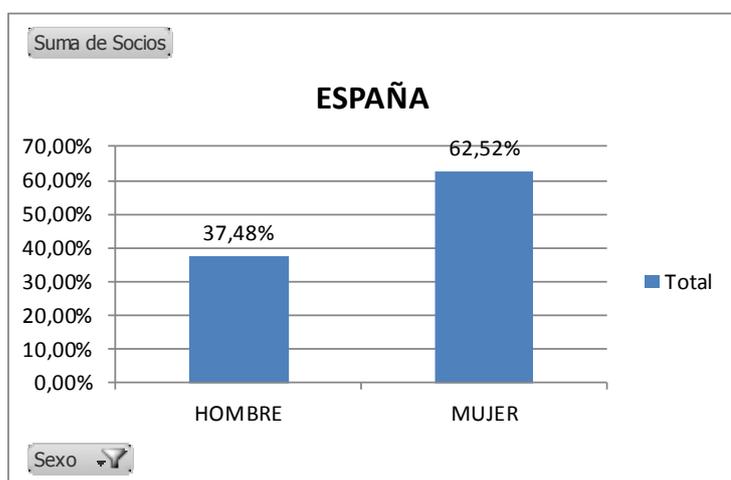
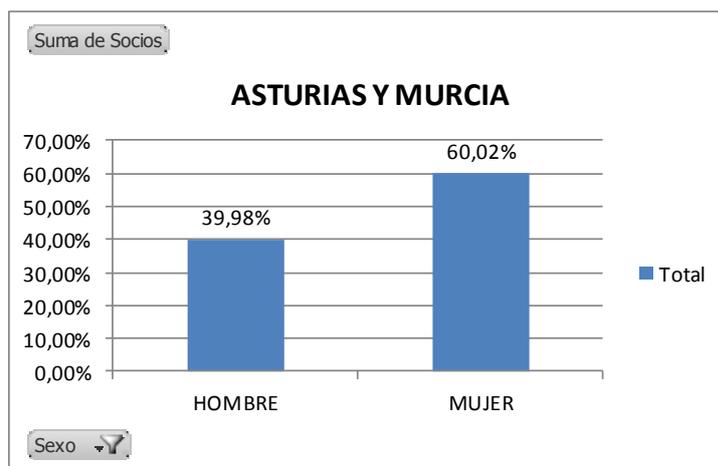
Región de rechazo de  $H_0$ :



**Región de rechazo de  $H_0$**  (Se amplía definición de Chi-Cuadrado en *Anexo 2*).

Variable SEXO:

Suma de Socios	
Sexo	Total
HOMBRE	39,98%
MUJER	60,02%
<b>Total general</b>	<b>100,00%</b>



Suma de Socios	
Sexo	Total
HOMBRE	37,48%
MUJER	62,52%
<b>Total general</b>	<b>100,00%</b>

(Figuras 11 y 12)

El histograma de frecuencias muestra aparentemente el mismo comportamiento de la variable sexo en la zona de test y en el conjunto de España, pero debido al gran volumen de datos, una diferencia en torno al 2% en esta variable, es muy grande. Se verificará este resultado a través del contraste de la Chi cuadrado.

A partir de las frecuencias observadas en la zona de Asturias y Murcia (ASMU), se calculan las frecuencias esperadas:

Sexo	ASMU	TOTAL	Frec_Observadas		Frec_Esperadas		Residual
			O <sub>i</sub>	E <sub>i</sub>	O <sub>i</sub> -E <sub>i</sub>		
HOMBRE	90.818	1.266.046	90.818	85.515	5.303		
MUJER	136.347	2.097.135	136.347	141.650	- 5.303		
	<b>227.165</b>	<b>3.363.181</b>					
			<b>227.165*1.266.046/3.363.181</b>				

(Figura 13)

En esta salida se muestra:

Frec. Observadas: es la frecuencia observada en la zona de Asturias y Murcia para cada sexo. Se puede ver que para el sexo hombre es superior a la esperada, y en el sexo mujer ocurre lo contrario.

Frec. Esperadas: es el valor esperado para cada sexo, calculado a partir de las frecuencias relativas de la zona de test.

Residual: muestra el residuo (frecuencia observada menos el valor esperado). La tabla muestra un residuo negativo para el sexo mujer, lo que significa que hay menos mujeres de las esperadas.

La siguiente tabla muestra el cálculo del contraste *Chi-cuadrado*:

$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$			
Sexo	O <sub>i</sub> -E <sub>i</sub>	(O <sub>i</sub> -E <sub>i</sub> ) <sup>2</sup>	(O <sub>i</sub> -E <sub>i</sub> ) <sup>2</sup> /E <sub>i</sub>
HOMBRE	5.303	28.125.226	329
MUJER	- 5.303	28.125.226	199
			<b>527</b>

(Figura 14)

Este valor hay que compararlo con el valor crítico de la distribución  $\chi^2$  con (2-1)=1 grados de libertad. Este valor es: 0.95 (1) 3.84. Puesto que el valor del estadístico (0) es mayor que el valor crítico, se debe rechazar la hipótesis nula, por tanto, se rechaza la hipótesis de independencia entre variables para el sexo y la zona.

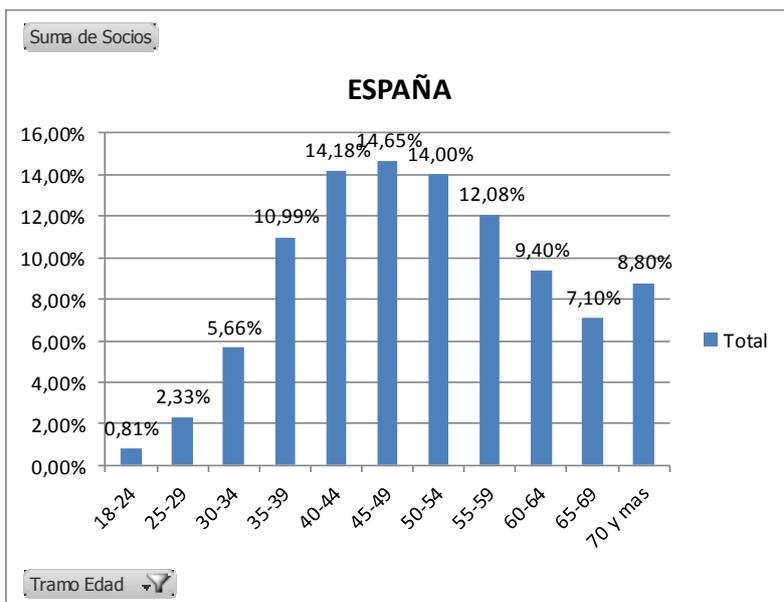
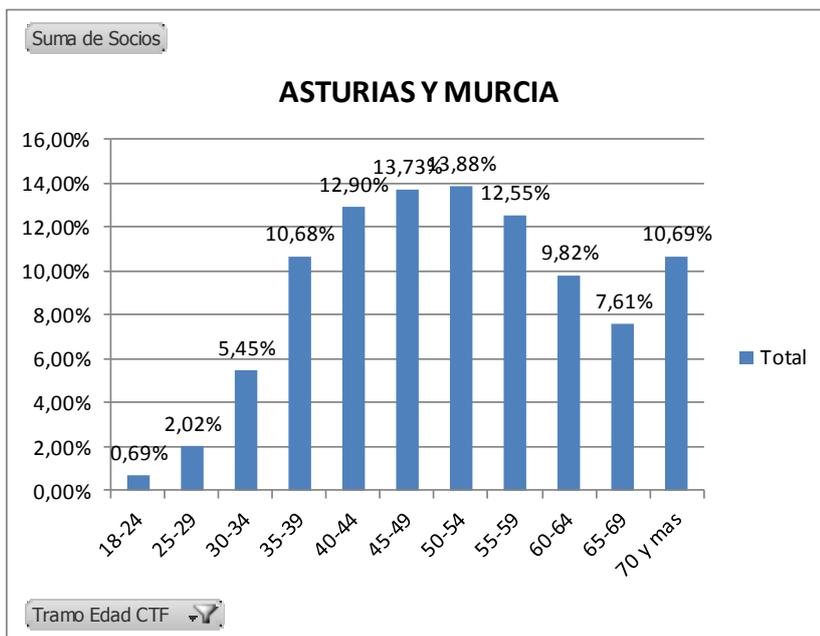
Estos resultados se han calculado también mediante el programa estadístico R, obteniéndose lo siguiente:

```
> x1 <- c(90818,136347)
> y1 <- c(1266046,2097135)
> chisq.test(x1, p=y1, rescale.p=TRUE)

Chi-squared test for given probabilities data: x1
X-squared = 527.4475, df = 1, p-value < 2.2e-16
```

Variable TRAMO EDAD:

Suma de Socios	
Tramo Edad C	Total
18-24	0,69%
25-29	2,02%
30-34	5,45%
35-39	10,68%
40-44	12,90%
45-49	13,73%
50-54	13,88%
55-59	12,55%
60-64	9,82%
65-69	7,61%
70 y mas	10,69%
<b>Total general</b>	<b>100,00%</b>



Suma de Socios	
Tramo Edad	Total
18-24	0,81%
25-29	2,33%
30-34	5,66%
35-39	10,99%
40-44	14,18%
45-49	14,65%
50-54	14,00%
55-59	12,08%
60-64	9,40%
65-69	7,10%
70 y mas	8,80%
<b>Total general</b>	<b>100,00%</b>

(Figuras 15 y 16)

Las distribuciones por Tramo\_Edad en ambas zonas a priori se comportan de la misma manera, a continuación se realiza el contraste de bondad de ajuste para comprobar si realmente existe independencia entre las dos variables.

Se toma como hipótesis nula que la variable Tramo\_Edad es independiente de la zona geográfica de la que se trate, es decir, la variable Tramo\_Edad se comporta igual en las regiones de Asturias y Murcia que en toda España.

Tabla de datos:

Tramo Edad	ASMU	TOTAL	Frec_Observadas	Frec_Esperadas	Residual
			O <sub>i</sub>	E <sub>i</sub>	O <sub>i</sub> -E <sub>i</sub>
18-24	1.480	25.950	1.480	1.712	- 232
25-29	4.332	75.133	4.332	4.956	- 624
30-34	11.695	183.730	11.695	12.120	- 425
35-39	22.916	357.001	22.916	23.550	- 634
40-44	27.693	458.461	27.693	30.243	- 2.550
45-49	29.459	474.719	29.459	31.315	- 1.856
50-54	29.781	455.208	29.781	30.028	- 247
55-59	26.928	394.001	26.928	25.990	938
60-64	21.071	306.773	21.071	20.236	835
65-69	16.322	232.222	16.322	15.319	1.003
70 y mas	22.943	290.324	22.943	19.151	3.792
	<b>214.620</b>	<b>3.253.522</b>			

(Figura 17)

En esta salida se muestra:

Frec\_Observadas: es la frecuencia observada en la zona de Asturias y Murcia para cada tramo. Se puede ver que hasta los 54 años, las frecuencias esperadas son superiores a las observadas.

Frec\_Esperadas: es el valor esperado para cada tramo de edad. Se calcula mediante una regla de tres en la que se tiene en cuenta el valor obtenido para cada tramo en el resto de España, aplicándolo al total de la muestra que se tiene en Asturias y Murcia.

Residual: muestra el residuo (frecuencia observada menos el valor esperado). La tabla muestra un residuo negativo para los tramos de edad hasta años, lo que significa que hay menos socios de los esperados en esos tramos.

La siguiente tabla muestra el cálculo del contraste *Chi-cuadrado*:

$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$			
Tramo Edad	O <sub>i</sub> -E <sub>i</sub>	(O <sub>i</sub> -E <sub>i</sub> ) <sup>2</sup>	(O <sub>i</sub> -E <sub>i</sub> ) <sup>2</sup> /E <sub>i</sub>
18-24	- 232	53.733	31
25-29	- 624	389.602	79
30-34	- 425	180.480	15
35-39	- 634	401.610	17
40-44	- 2.550	6.500.359	215
45-49	- 1.856	3.444.908	110
50-54	- 247	61.006	2
55-59	938	879.004	34
60-64	835	696.532	34
65-69	1.003	1.006.760	66
70 y mas	3.792	14.376.614	751
			<b>1.354</b>

(Figura 18)

Este valor se debe comparar con el valor crítico de la distribución  $\chi^2$  con  $(11-1)=10$  grados de libertad. Este valor es: 0.95 (10) 18.31. Puesto que el valor del estadístico (1.354) es mayor que el valor crítico, de nuevo se rechaza la hipótesis nula, por tanto, se rechaza la hipótesis de independencia entre variables para el Tramo\_Edad y la zona.

Estos resultados se han calculado también mediante el programa estadístico R, obteniéndose lo siguiente:

```
> x2 <-c(1480,4332,11695,22916,27693,29459,29781,26928,21071,16322,22943)
> y2 <-c(25950,75133,183730,357001,458461,474719,455208,394001,306773,232222,290324)
> chisq.test(x2, p=y2, rescale.p=TRUE)

Chi-squared test for given probabilities data: x2

X-squared = 1353.57, df = 10, p-value < 2.2e-16
```

#### Variable SEGMENTO:

La variable segmento identifica a los socios en función de su comportamiento transaccional y se calcula teniendo en cuenta la compra, junto con la recencia y antigüedad del socio. De esta manera, cada socio pertenece a uno de estos segmentos:

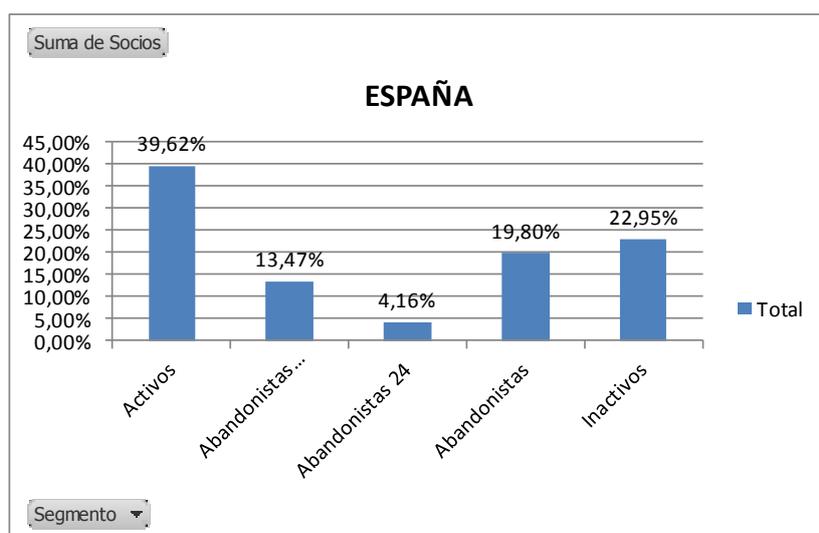
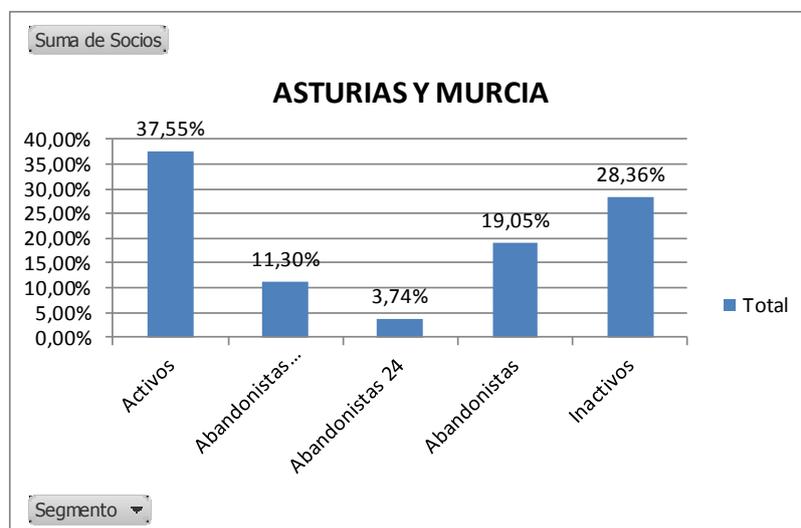
Activo 0-6	Comprador en el último año con antigüedad de 0 a 6 meses
Activo 7-12	Comprador en el último año con antigüedad de 7 a 12 meses
Activo x años	Comprador en el último año con antigüedad de x años
Abandonista Reciente x años	Comprador hace 2 años con antigüedad de x años
Abandonista 24, x años	Comprador hace 3 años con antigüedad de x años
Abandonista x años	Comprador hace 4, 5 o más años (sin antigüedad)
Inactivo 0-6	No comprador con antigüedad de 0 a 6 meses
Inactivo 7-12	No comprador con antigüedad de 7 a 12 meses
Inactivo x años	No comprador con antigüedad x años

Nota: x puede tomar los valores de 2, 3, 4, 5 y +5 años

(Figura 19)

Se obtienen los histogramas de frecuencias:

Suma de Socios	
Segmento	Total
Activos	37,55%
Abandonistas Recientes	11,30%
Abandonistas 24	3,74%
Abandonistas	19,05%
Inactivos	28,36%
<b>Total general</b>	<b>100,00%</b>



Suma de Socios	
Segmento	Total
Activos	39,62%
Abandonistas Re	13,47%
Abandonistas 24	4,16%
Abandonistas	19,80%
Inactivos	22,95%
<b>Total general</b>	<b>100,00%</b>

(Figuras 20 y 21)

Se observan proporciones similares para todos los segmentos, variando mínimamente el de activos e inactivos, pero hay que tener en cuenta el gran volumen de datos que se tiene para el conjunto de España, por lo que las pequeñas diferencias sí pueden ser muy significativas. Se calcula el contraste de Chi cuadrado para verificarlo.

En el contraste de bondad de ajuste, se toma como hipótesis nula que la variable Segmento es independiente de la zona geográfica de la que se trate, es decir, se comporta igual en las regiones de Asturias y Murcia que en el total de España.

Tabla de datos:

Segmento	ASMU	TOTAL	Frec_Observadas	Frec_Esperadas	Residual
			O <sub>i</sub>	E <sub>i</sub>	O <sub>i</sub> -E <sub>i</sub>
Activos	85.346	1.328.901	85.346	96.219	- 10.873
Abandonistas Recientes	25.671	448.502	25.671	32.474	- 6.803
Abandonistas 24	8.511	139.037	8.511	10.067	- 1.556
Abandonistas	43.291	664.878	43.291	48.140	- 4.849
Inactivos	64.453	784.856	64.453	56.827	7.626
	<b>227.272</b>	<b>3.138.902</b>			

(Figura 22)

En esta salida se muestra:

Frec. Observadas: es la frecuencia observada en la zona de Asturias y Murcia para cada segmento. Se puede ver que en los segmentos de activos y abandonistas, las frecuencias esperadas son superiores a las observadas.

Frec. Esperadas: es el valor esperado para cada segmento. Se calcula mediante una regla de tres en la que se tiene en cuenta el valor obtenido para cada tramo en el resto de España, aplicándolo al total de la muestra que se tiene en Asturias y Murcia. Solamente es superior al valor observado el segmento de inactivos.

Residual: muestra el residuo (frecuencia observada menos el valor esperado). La tabla muestra un residuo positivo solamente para el segmento de inactivos, lo que significa que hay más socios de los esperados.

La siguiente tabla muestra el cálculo del contraste *Chi-cuadrado*

$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$			
Segmento	O <sub>i</sub> -E <sub>i</sub>	(O <sub>i</sub> -E <sub>i</sub> ) <sup>2</sup>	(O <sub>i</sub> -E <sub>i</sub> ) <sup>2</sup> /E <sub>i</sub>
Activos	- 10.873	118.221.966	1.229
Abandonistas Recientes	- 6.803	46.277.528	1.425
Abandonistas 24	- 1.556	2.421.028	240
Abandonistas	- 4.849	23.517.159	489
Inactivos	7.626	58.149.030	1.023
			<b>4.406</b>

(Figura 23)

Este valor debe ser compararlo con el valor crítico de la distribución  $\chi^2$  con  $(5-1)=4$  grados de libertad. Este valor es: 0.95 (4) 9.49. Puesto que el valor del estadístico (4.406) es mayor que el valor crítico, se debe rechazar la hipótesis nula, por tanto, se rechaza la hipótesis de independencia entre variables para el Segmento y la zona.

Estos resultados se han calculado también mediante el programa estadístico R, obteniéndose lo siguiente:

```
> x3 <- c(85346,25671,8511,43291,64453)
> y3 <- c(1328901,448502,139037,664878,784856)
> chisq.test(x3, p=y3, rescale.p=TRUE)

Chi-squared test for given probabilities data: x3

X-squared = 4405.559, df = 4, p-value < 2.2e-16
```

o Variables económicas:

De la misma manera, se analiza el comportamiento de los socios de ambas zonas en función del valor de su compra. A partir de diferentes métricas obtenidas directamente de la base de datos, se calculan las variables que se van estudiar.

Se definen las tres variables:

Venta Neta ~ VN: importe de las compras realizadas por el socio.

Operaciones ~ Op.: número de veces que el socio ha realizado una compra.

Ticket Medio ~ TM = Venta/Nº Operaciones

Media de Compra ~ MdC = Venta/Nº Compradores

Artículos por Operación ~ Art x Op = Nº Artículos/Nº Operaciones

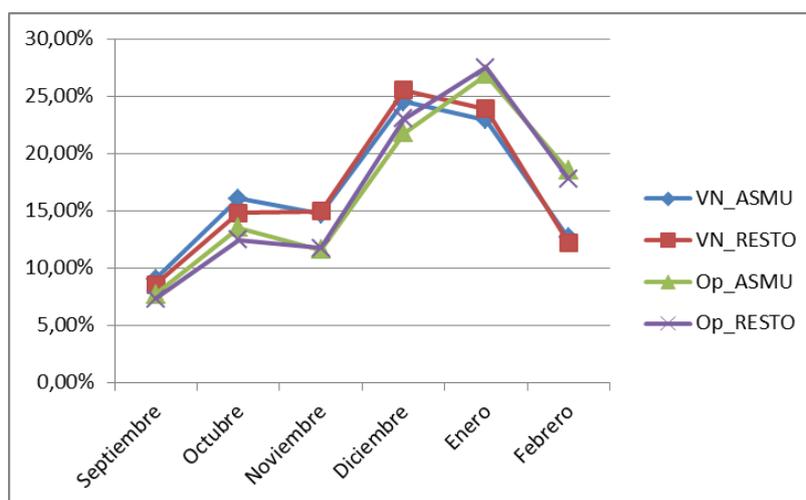
Frecuencia de Compra ~ Frec = Nº Operaciones/Nº Compradores

Se extrae información referente a la temporada otoño-invierno, que se corresponde con los meses de septiembre a febrero mediante un informe de MicroStrategy y se obtienen los histogramas de frecuencias para observar el resultado.

Se observan, por un lado, Venta Neta y Operaciones, Ticket Medio y Media de Compra, y Artículos por Operación y Frecuencia de Compra. Y por otro lado, se calculan las principales medidas de tendencia central y de dispersión de las variables VN, Operaciones y Ticket Medio.

Variables VENTA NETA y OPERACIONES:

Mes Año	VN_ASMU	VN_RESTO	Op_ASMU	Op_RESTO
Septiembre	9,07%	8,60%	7,71%	7,35%
Octubre	16,07%	14,81%	13,52%	12,49%
Noviembre	14,73%	14,98%	11,60%	11,78%
Diciembre	24,53%	25,51%	21,76%	23,06%
Enero	22,92%	23,89%	26,87%	27,50%
Febrero	12,69%	12,22%	18,53%	17,82%
<b>Total general</b>	<b>100,00%</b>	<b>100,00%</b>	<b>100,00%</b>	<b>100,00%</b>

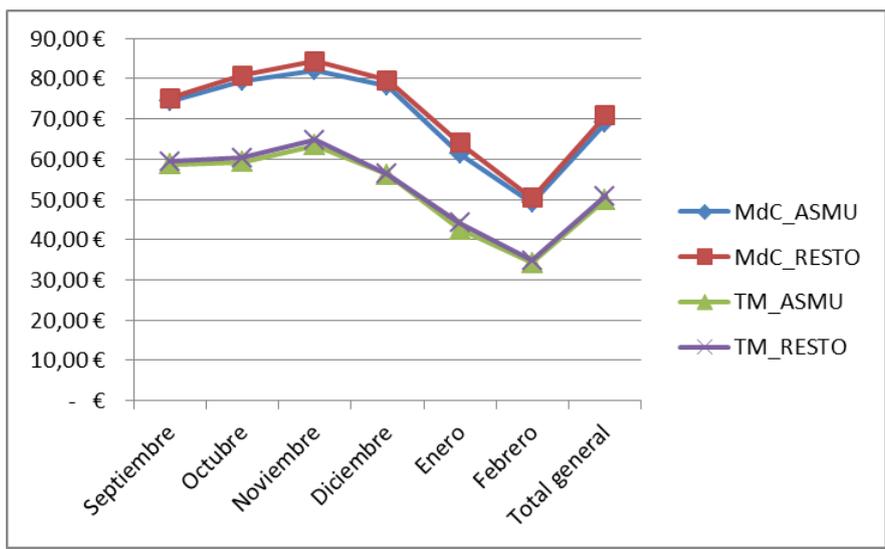


(Figura 24)

Se observan variaciones a principio de temporada y también a principios de año.

Variables MEDIA DE COMPRA y TICKET MEDIO:

Mes Año	MdC_ASMU	MdC_RESTO	TM_ASMU	TM_RESTO
Septiembre	74,41 €	75,21 €	58,69 €	59,49 €
Octubre	79,44 €	80,77 €	59,27 €	60,33 €
Noviembre	82,14 €	84,26 €	63,36 €	64,70 €
Diciembre	78,28 €	79,56 €	56,22 €	56,28 €
Enero	61,28 €	64,08 €	42,54 €	44,19 €
Febrero	49,13 €	50,31 €	34,16 €	34,89 €
<b>Total general</b>	<b>69,01 €</b>	<b>70,84 €</b>	<b>49,89 €</b>	<b>50,88 €</b>

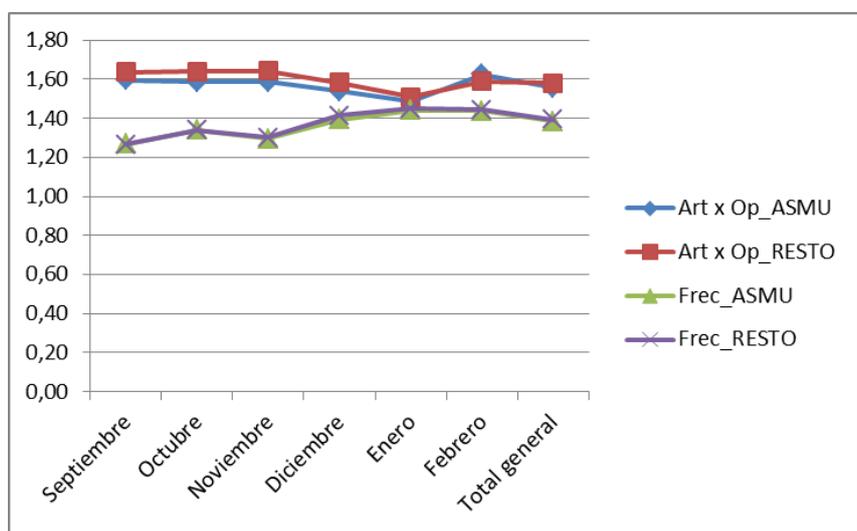


(Figura 25)

Se puede ver que apenas existe variación de la variable Ticket Medio y una mínima variación en la Media de Compra.

Variables ARTÍCULOS POR OPERACIÓN y FRECUENCIA DE COMPRA:

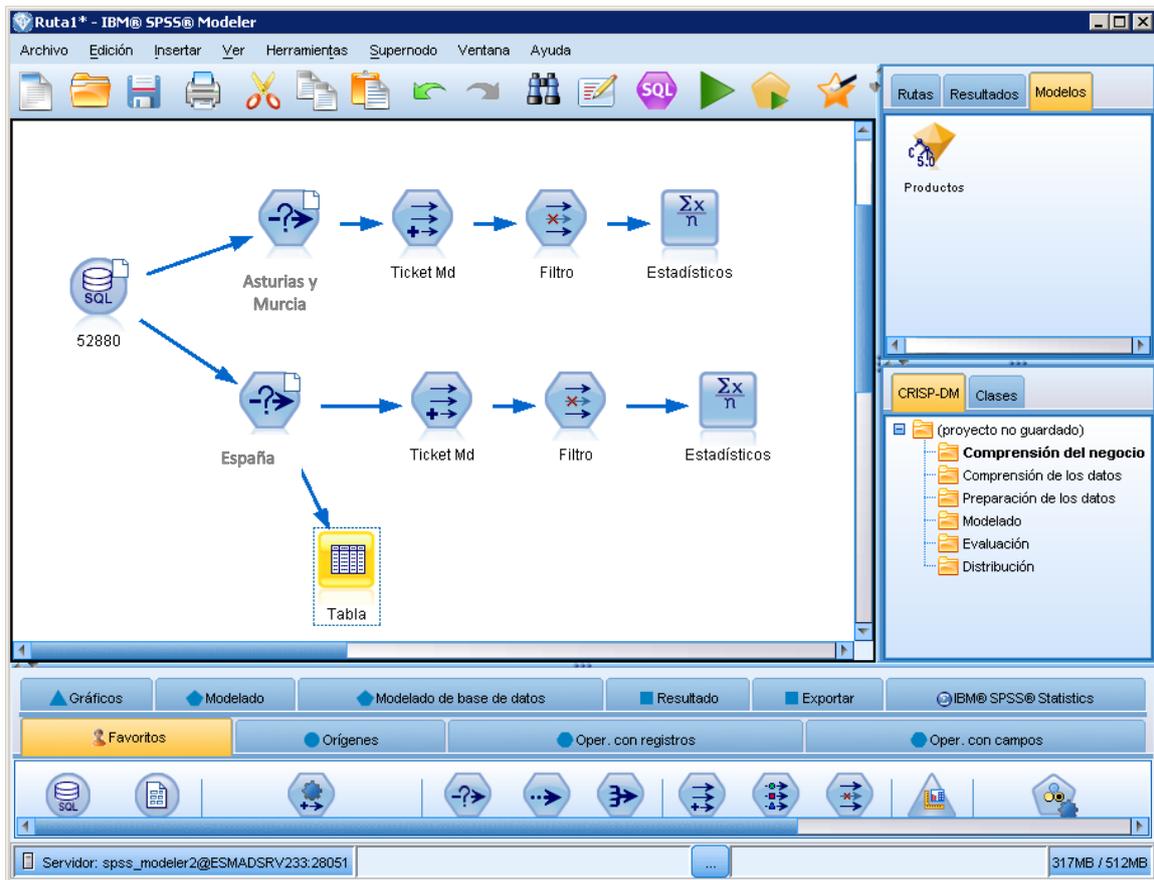
Mes Año	Art x Op_ASMU	Art x Op_RESTO	Frec_ASMU	Frec_RESTO
Septiembre	1,59	1,64	1,3	1,3
Octubre	1,58	1,64	1,3	1,3
Noviembre	1,58	1,64	1,3	1,3
Diciembre	1,54	1,58	1,4	1,4
Enero	1,48	1,51	1,4	1,5
Febrero	1,63	1,59	1,4	1,4
<b>Total general</b>	<b>1,56</b>	<b>1,58</b>	<b>1,4</b>	<b>1,4</b>



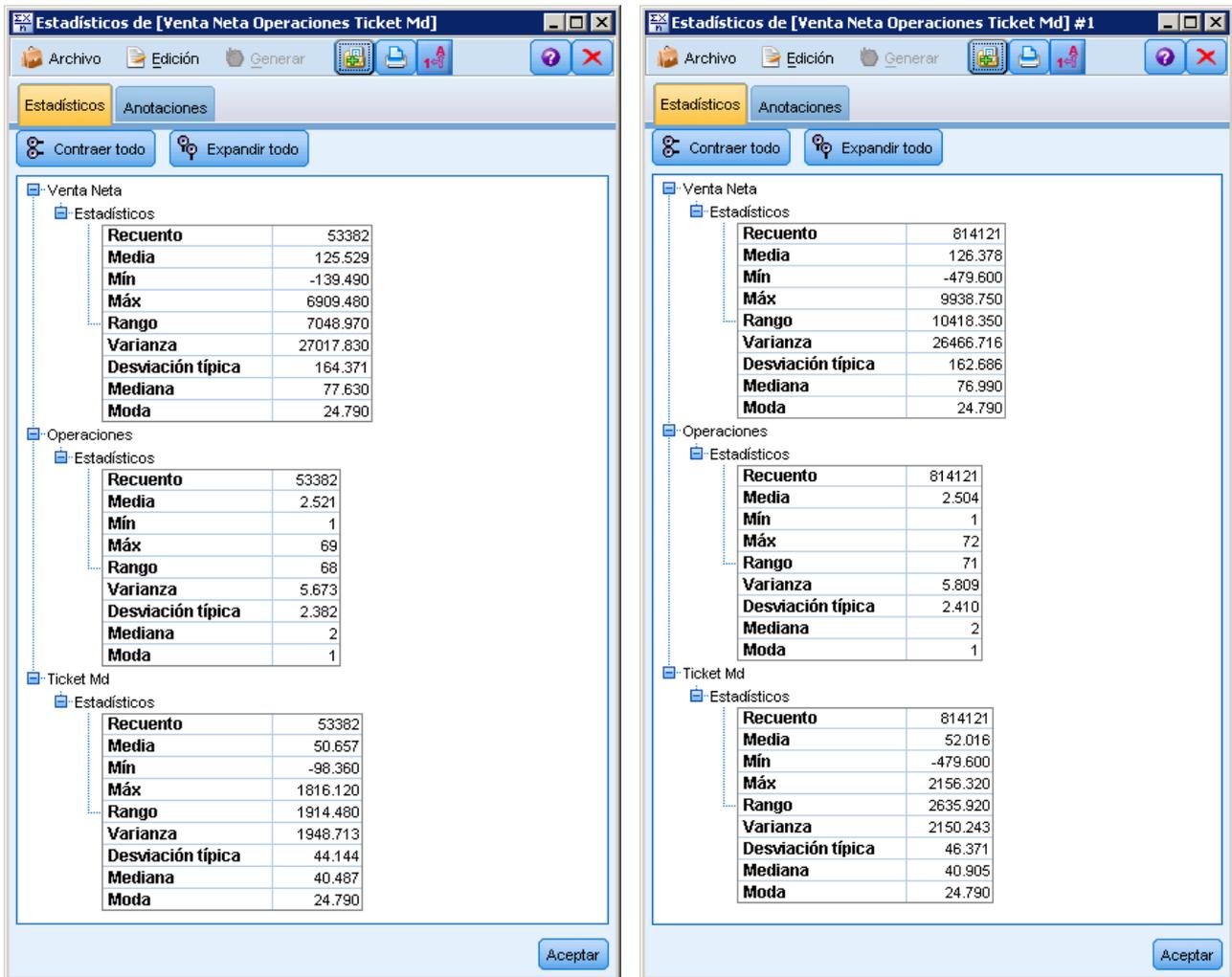
(Figura 26)

Se observa que el número de Artículos por Operación varía mínimamente y la frecuencia es prácticamente la misma.

Mediante SPSS Modeler, se calculan además, las principales medidas de tendencia central y de dispersión de las variables VN, Operaciones y Ticket Medio, en Asturias y Murcia y en el conjunto de España, con el fin de estudiarlas también:



(Figura 27)



(Figura 28)

### 5.3. Comportamiento de compra

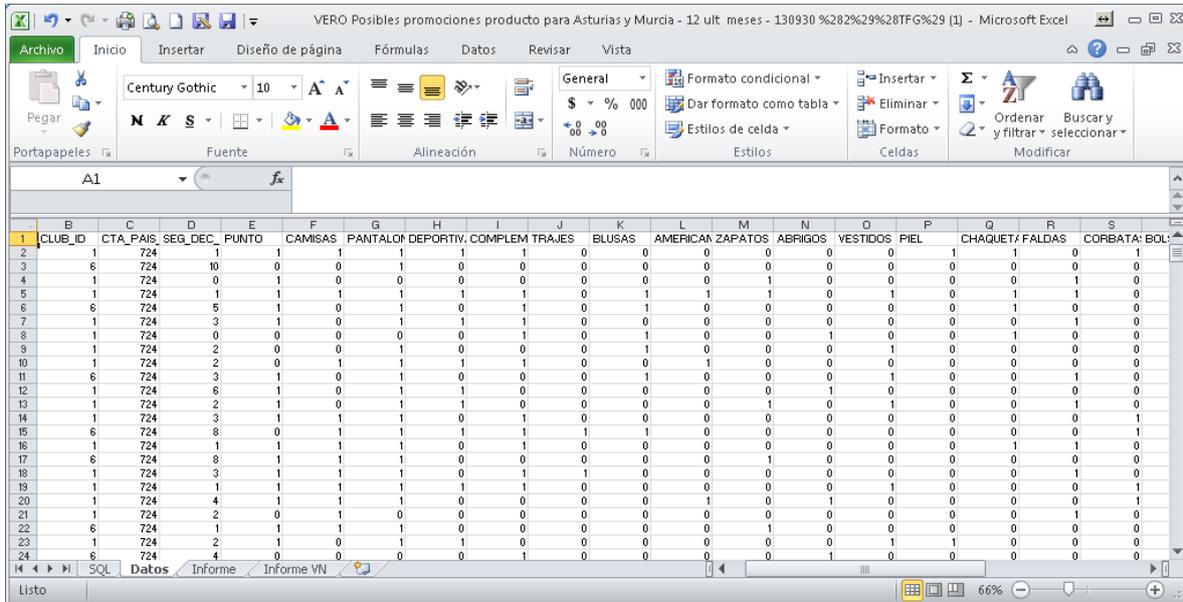
En este apartado se mostrarán las decisiones tomadas por la dirección comercial, en base a los informes entregados referentes a las familias de producto en las que los socios realizan sus compras. A continuación se realizará un modelo con el que se concluirá si las decisiones tomadas fueron las apropiadas.

- Selección de las familias de producto por parte de la Dirección

Las categorías de producto que se utilizarán no están calculadas dentro del CRM, por lo que se plantea una sentencia de programación mediante la herramienta SQL, que ataca directamente a las tablas originales.

Las fechas que se han tomado son del 1 de septiembre de 2012 al 28 de febrero de 2013, periodo que comprende una temporada comercial cerrada. Se seleccionan las zonas de Asturias y Murcia. Para cada socio, se calculan determinadas variables económicas junto con otras variables que se crean en función de las prendas que ha comprado en el periodo dado.

Como resultado se obtiene un listado de socios, cada uno de los cuales viene marcado con 1 si han comprado y 0 si no han comprado en cada una de las familias:



(Figura 29)

Se muestra la primera fila de los datos obtenidos, para ver un ejemplo de lo que se tiene:

CLUB_ID	CTA_PAIS_ID	SEG_DEC_A	PUNTO	CAMISAS	PANTALON	DEPORTIVAS	COMPLEMEN	TRAJES	BLUSAS
1	724	1	1	1	1	1	1	0	0

AMERICANA	ZAPATOS	ABRIGOS	VESTIDOS	PIEL	CHAQUETAS	FALDAS	CORBATAS	BOLSOS	SEXO_ID
0	0	0	0	1	1	0	1	0	H

PANTALON+FALDA	CAMISAS+BLUSAS
1	1

Para este socio en concreto, se observa que ha comprado en las secciones de punto, camisas, pantalón, deportivas, complementos, piel, chaquetas y corbatas, pero no ha comprado en las de trajes, blusas, americanas, zapatos, abrigos, vestidos, faldas ni bolsos.

Para poder trabajar mejor, se crean las agrupaciones pantalón+falda (también se nombrará como parte de abajo) y camisas+blusas (también se nombrará como parte de arriba); esto se debe a que el pantalón y la falda son partes de abajo genéricas cuya única diferencia es la sección en la que se venden, caballero o señora. Lo mismo ocurre para las camisas y blusas, pero en este caso ambas son partes de arriba.

La información obtenida se puede ver ordenada tras su manipulación en un fichero Excel a través de diferentes tablas dinámicas, tantas como familias se tienen. Así, para cada familia se crea una tabla en la que se incluye como filtro el resto de familias, marcándolas como “sin compra”, de manera que vamos a detectar cuáles son las preferidas por nuestros socios:

PUNTO	Sin punto	↕
CAMISAS+BLUSAS	Sin camisa/blusa	↕
PANTALON+FALDA	Sin pantalon/falda	↕
DEPORTIVAS	Sin deportivas	↕
COMPLEMENTOS	Sin complementos	↕
TRAJES	Sin Trajes	↕
AMERICANAS	Sin Americanas	↕
PIEL	Sin Piel	↕
CORBATAS	Sin Corbatas	↕
BOLSOS	Sin Bolsos	↕
CHAQUETAS	Sin chaquetas	↕
ZAPATOS	Sin zapatos	↕
VESTIDOS	Sin vestidos	↕

Venta Neta		Total
ABRIGOS	▼	
Sin abrigos		5.742 €
Con abrigos		30.339 €
<b>Total general</b>		<b>36.081 €</b>

PUNTO	Sin punto	↕
CAMISAS+BLUSAS	Sin camisa/blusa	↕
PANTALON+FALDA	Sin pantalon/falda	↕
DEPORTIVAS	Sin deportivas	↕
COMPLEMENTOS	Sin complementos	↕
TRAJES	Sin Trajes	↕
AMERICANAS	Sin Americanas	↕
PIEL	Sin Piel	↕
CORBATAS	Sin Corbatas	↕
BOLSOS	Sin Bolsos	↕
CHAQUETAS	Sin chaquetas	↕
ZAPATOS	Sin zapatos	↕
ABRIGOS	Sin abrigos	↕

Venta Neta		Total
VESTIDOS	▼	
Sin vestidos		5.742 €
Con vestidos		40.292 €
<b>Total general</b>		<b>46.034 €</b>

0,3%

0,4%

PUNTO	Sin punto	↕
CAMISAS+BLUSAS	Sin camisa/blusa	↕
PANTALON+FALDA	Sin pantalon/falda	↕
DEPORTIVAS	Sin deportivas	↕
COMPLEMENTOS	Sin complementos	↕
TRAJES	Sin Trajes	↕
AMERICANAS	Sin Americanas	↕
PIEL	Sin Piel	↕
CORBATAS	Sin Corbatas	↕
BOLSOS	Sin Bolsos	↕
CHAQUETAS	Sin chaquetas	↕
VESTIDOS	Sin vestidos	↕
ABRIGOS	Sin abrigos	↕

Venta Neta		Total
ZAPATOS	▼	
Sin zapatos		5.742 €
Con zapatos		16.112 €
<b>Total general</b>		<b>21.854 €</b>

0,2%

PUNTO	Sin punto	↕
CAMISAS+BLUSAS	Sin camisa/blusa	↕
PANTALON+FALDA	Sin pantalon/falda	↕
DEPORTIVAS	Sin deportivas	↕
COMPLEMENTOS	Sin complementos	↕
TRAJES	Sin Trajes	↕
AMERICANAS	Sin Americanas	↕
PIEL	Sin Piel	↕
CORBATAS	Sin Corbatas	↕
BOLSOS	Sin Bolsos	↕
ZAPATOS	Sin zapatos	↕
VESTIDOS	Sin vestidos	↕
ABRIGOS	Sin abrigos	↕

Venta Neta		Total
CHAQUETAS	▼	
Sin chaquetas		5.742 €
Con chaquetas		16.002 €
<b>Total general</b>		<b>21.744 €</b>

PUNTO	Sin punto	↕
CAMISAS+BLUSAS	Sin camisa/blusa	↕
PANTALON+FALDA	Sin pantalon/falda	↕
DEPORTIVAS	Sin deportivas	↕
COMPLEMENTOS	Sin complementos	↕
TRAJES	Sin Trajes	↕
AMERICANAS	Sin Americanas	↕
PIEL	Sin Piel	↕
CORBATAS	Sin Corbatas	↕
CHAQUETAS	Sin chaquetas	↕
ZAPATOS	Sin zapatos	↕
VESTIDOS	Sin vestidos	↕
ABRIGOS	Sin abrigos	↕

Venta Neta		Total
BOLSOS	▼	
Sin Bolsos		5.742 €
Con bolsos		3.569 €
<b>Total general</b>		<b>9.311 €</b>

0,1%

0,0%

PUNTO	Sin punto	↕
CAMISAS+BLUSAS	Sin camisa/blusa	↕
PANTALON+FALDA	Sin pantalon/falda	↕
DEPORTIVAS	Sin deportivas	↕
COMPLEMENTOS	Sin complementos	↕
TRAJES	Sin Trajes	↕
AMERICANAS	Sin Americanas	↕
PIEL	Sin Piel	↕
BOLSOS	Sin Bolsos	↕
CHAQUETAS	Sin chaquetas	↕
ZAPATOS	Sin zapatos	↕
VESTIDOS	Sin vestidos	↕
ABRIGOS	Sin abrigos	↕

Venta Neta		Total
CORBATAS	▼	
Sin corbatas		5.742 €
Con corbatas		4.200 €
<b>Total general</b>		<b>9.942 €</b>

0,0%

PUNTO	Sin punto	↕
CAMISAS+BLUSAS	Sin camisa/blusa	↕
PANTALON+FALDA	Sin pantalon/falda	↕
DEPORTIVAS	Sin deportivas	↕
COMPLEMENTOS	Sin complementos	↕
TRAJES	Sin Trajes	↕
AMERICANAS	Sin Americanas	↕
CORBATAS	Sin Corbatas	↕
BOLSOS	Sin Bolsos	↕
CHAQUETAS	Sin chaquetas	↕
ZAPATOS	Sin zapatos	↕
VESTIDOS	Sin vestidos	↕
ABRIGOS	Sin abrigos	↕

Venta Neta		Total
PIEL	▼	
Sin Piel		5.742 €
Con piel		19.629 €
<b>Total general</b>		<b>25.371 €</b>

PUNTO	Sin punto	↕
CAMISAS+BLUSAS	Sin camisa/blusa	↕
PANTALON+FALDA	Sin pantalon/falda	↕
DEPORTIVAS	Sin deportivas	↕
COMPLEMENTOS	Sin complementos	↕
TRAJES	Sin Trajes	↕
PIEL	Sin Piel	↕
CORBATAS	Sin Corbatas	↕
BOLSOS	Sin Bolsos	↕
CHAQUETAS	Sin chaquetas	↕
ZAPATOS	Sin zapatos	↕
VESTIDOS	Sin vestidos	↕
ABRIGOS	Sin abrigos	↕

Venta Neta		Total
AMERICANAS	▼	
Sin Americanas		5.742 €
Con Americanas		9.895 €
<b>Total general</b>		<b>15.637 €</b>

0,2%

0,1%

PUNTO	Sin punto	↕
CAMISAS+BLUSAS	Sin camisa/blusa	↕
PANTALON+FALDA	Sin pantalon/falda	↕
DEPORTIVAS	Sin deportivas	↕
COMPLEMENTOS	Sin complementos	↕
TRAJES	Sin Trajes	↕
AMERICANAS	Sin Americanas	↕
CORBATAS	Sin Corbatas	↕
BOLSOS	Sin Bolsos	↕
CHAQUETAS	Sin chaquetas	↕
ZAPATOS	Sin zapatos	↕
VESTIDOS	Sin vestidos	↕
ABRIGOS	Sin abrigos	↕

Venta Neta		Total
PIEL	▼	
Sin Piel		5.742 €
Con piel		19.629 €
<b>Total general</b>		<b>25.371 €</b>

0,2%

PUNTO	Sin punto				
CAMISAS+BLUSAS	Sin camisa/blusa				
PANTALON+FALDA	Sin pantalon/falda				
DEPORTIVAS	Sin deportivas				
TRAJES	Sin Trajes				
AMERICANAS	Sin Americanas				
PIEL	Sin Piel				
CORBATAS	Sin Corbatas				
BOLSOS	Sin Bolsos				
CHAQUETAS	Sin chaquetas				
ZAPATOS	Sin zapatos				
VESTIDOS	Sin vestidos				
ABRIGOS	Sin abrigos				

Venta Neta		Total			
COMPLEMENTOS					
Sin complementos		5.742 €			
Con complementos		41.261 €	0,3%		
<b>Total general</b>		<b>47.003 €</b>			

PUNTO	Sin punto				
CAMISAS+BLUSAS	Sin camisa/blusa				
PANTALON+FALDA	Sin pantalon/falda				
DEPORTIVAS	Sin deportivas				
COMPLEMENTOS	Sin complementos				
TRAJES	Sin Trajes				
AMERICANAS	Sin Americanas				
PIEL	Sin Piel				
CORBATAS	Sin Corbatas				
BOLSOS	Sin Bolsos				
CHAQUETAS	Sin chaquetas				
ZAPATOS	Sin zapatos				
VESTIDOS	Sin vestidos				
ABRIGOS	Sin abrigos				

Venta Neta		Total			
DEPORTIVAS					
Sin deportivas		5.742 €			
Con deportivas		70.786 €	0,6%		
<b>Total general</b>		<b>76.528 €</b>			

PUNTO	Sin punto				
CAMISAS+BLUSAS	Sin camisa/blusa				
PANTALON+FALDA	Sin pantalon/falda				
DEPORTIVAS	Sin deportivas				
COMPLEMENTOS	Sin complementos				
TRAJES	Sin Trajes				
AMERICANAS	Sin Americanas				
PIEL	Sin Piel				
CORBATAS	Sin Corbatas				
BOLSOS	Sin Bolsos				
CHAQUETAS	Sin chaquetas				
ZAPATOS	Sin zapatos				
VESTIDOS	Sin vestidos				
ABRIGOS	Sin abrigos				

Venta Neta		Total			
PANTALON+FALDA					
Sin pantalon/falda		5.742 €			
Con pantalon/falda		162.249 €	1,3%		
<b>Total general</b>		<b>167.991 €</b>			

PUNTO	Sin punto				
PANTALON+FALDA	Sin pantalon/falda				
DEPORTIVAS	Sin deportivas				
COMPLEMENTOS	Sin complementos				
TRAJES	Sin Trajes				
AMERICANAS	Sin Americanas				
PIEL	Sin Piel				
CORBATAS	Sin Corbatas				
BOLSOS	Sin Bolsos				
CHAQUETAS	Sin chaquetas				
ZAPATOS	Sin zapatos				
VESTIDOS	Sin vestidos				
ABRIGOS	Sin abrigos				

Venta Neta		Total			
CAMISAS+BLUSAS					
Sin camisa/blusa		5.742 €			
Con camisa/blusa		129.515 €	1,1%		
<b>Total general</b>		<b>135.257 €</b>			

PUNTO	Sin punto				
CAMISAS+BLUSAS	Sin camisa/blusa				
PANTALON+FALDA	Sin pantalon/falda				
DEPORTIVAS	Sin deportivas				
COMPLEMENTOS	Sin complementos				
TRAJES	Sin Trajes				
AMERICANAS	Sin Americanas				
PIEL	Sin Piel				
CORBATAS	Sin Corbatas				
BOLSOS	Sin Bolsos				
CHAQUETAS	Sin chaquetas				
ZAPATOS	Sin zapatos				
VESTIDOS	Sin vestidos				
ABRIGOS	Sin abrigos				

Venta Neta		Total			
PUNTO					
Sin punto		5.742 €			
Con punto		372.032 €	3,0%		
<b>Total general</b>		<b>377.774 €</b>			

(Figura 30)

Se pretende extraer compradores únicos para cada familia, de manera que se seleccionaran aquellas cinco en las que el porcentaje de venta sea mayor. Se limitan a cinco familias por decisión de la dirección en base a cuestiones económicas.

Por otro lado, se sabe que si se combinan las diferentes familias ya elegidas, el porcentaje de socios compradores es mayor; de esta forma se conseguirá un número de socios suficiente en cuanto a gasto de generación de la campaña y apropiado para la obtención de resultados. Se calculará mediante otra tabla dinámica, en la que se realizan todas las combinaciones posibles de las cinco familias seleccionadas.

Los resultados se presentan a la Dirección en una reunión y, a partir de la venta generada por cada agrupación de familias, junto con las posibilidades que ofrecen los departamentos de ventas y marketing en base a resultados propios de su área, se eligen las categorías para las que se segmentará a los socios.

Tabla de Producto para las cinco familias elegidas

PUNTO	CAMISAS+BLUSAS	PANTALON+FALDA	DEPORTIVAS	COMPLEMENTOS	Nº socios	% socios s/ total		
Sin punto	Sin camisa/blusa	Sin pantalon/falda	Sin deportivas	Sin complementos	2.992	4%		
			Con deportivas	Con complementos	2.105	3%		
		Con pantalon/falda	Sin deportivas	Sin complementos	1.260	2%		
			Con deportivas	Con complementos	151	0%		
		Con camisa/blusa	Sin pantalon/falda	Sin deportivas	Sin complementos	4.930	7%	
				Con deportivas	Con complementos	729	1%	
	Con pantalon/falda		Sin deportivas	Sin complementos	354	0%		
			Con deportivas	Con complementos	78	0%		
	Con punto		Sin camisa/blusa	Sin pantalon/falda	Sin deportivas	Sin complementos	4.644	6%
					Con deportivas	Con complementos	849	1%
		Con pantalon/falda	Sin deportivas	Sin complementos	357	0%		
			Con deportivas	Con complementos	104	0%		
Con punto		Sin camisa/blusa	Sin pantalon/falda	Sin deportivas	Sin complementos	2.389	3%	
				Con deportivas	Con complementos	752	1%	
	Con pantalon/falda	Sin deportivas	Sin complementos	315	0%			
		Con deportivas	Con complementos	161	0%			
	Con punto	Sin camisa/blusa	Sin pantalon/falda	Sin deportivas	Sin complementos	11.945	16%	
				Con deportivas	Con complementos	2.173	3%	
Con pantalon/falda			Sin deportivas	Sin complementos	886	1%		
			Con deportivas	Con complementos	295	0%		
Con camisa/blusa			Sin pantalon/falda	Sin deportivas	Sin complementos	6.239	8%	
				Con deportivas	Con complementos	2.370	3%	
		Con pantalon/falda	Sin deportivas	Sin complementos	866	1%		
			Con deportivas	Con complementos	593	1%		
		Con punto	Sin camisa/blusa	Sin pantalon/falda	Sin deportivas	Sin complementos	5.256	7%
					Con deportivas	Con complementos	1.872	3%
Con pantalon/falda			Sin deportivas	Sin complementos	734	1%		
			Con deportivas	Con complementos	430	1%		
Con punto	Sin camisa/blusa		Sin pantalon/falda	Sin deportivas	Sin complementos	7.208	10%	
				Con deportivas	Con complementos	5.831	8%	
Con punto	Sin camisa/blusa	Sin pantalon/falda	Sin deportivas	Sin complementos	1.976	3%		
			Con deportivas	Con complementos	2.841	4%		
<b>Total general</b>					<b>73.685</b>	<b>100%</b>		

(Figura 31)

De esta manera, para la segmentación se seleccionan 13 versiones diferentes de la promoción con un incentivo añadido que se enviará a cada grupo de socios que compran con las combinaciones de familias elegidas, más la versión general que se envía al resto de socios. La selección debe ser limitada en cuanto al número de socios para minimizar gastos.

En resumen, estas serán las combinaciones que se tienen para realizar la segmentación:

1	2.105	3%	Solo complementos
2	1.260	2%	Solo deportivas
3	4.930	7%	Solo pantalon/falda
4	4.644	6%	Solo camisa/blusa
5	2.389	3%	Camisa/blusa+pantalon/falda
6	11.945	16%	Solo punto
7	2.173	3%	Punto + complemento
8	6.239	8%	Punto + pantalon/falda
9	2.370	3%	Punto + pantalon/falda + complemento
10	5.256	7%	Punto + camisa/blusa
11	1.872	3%	Punto + camisa/blusa + complemento
12	7.208	10%	Punto + camisa/blusa + pantalon/falda
13	5.831	8%	Punto + camisa/blusa + pantalon/falda + complemento

(Figura 32)

- o Selección de las familias de producto

Desde el punto de vista estadístico se puede identificar grupos de socios utilizando herramientas analíticas y de modelado. En este caso se dispone de la herramienta de minería de datos denominada de SPSS Modeler.

De todos los modelos que dispone la herramienta, se va a optar por realizar un modelo denominado “A Priori”, que sirve para buscar reglas de asociaciones, en este caso, entre los productos que compran los socios.

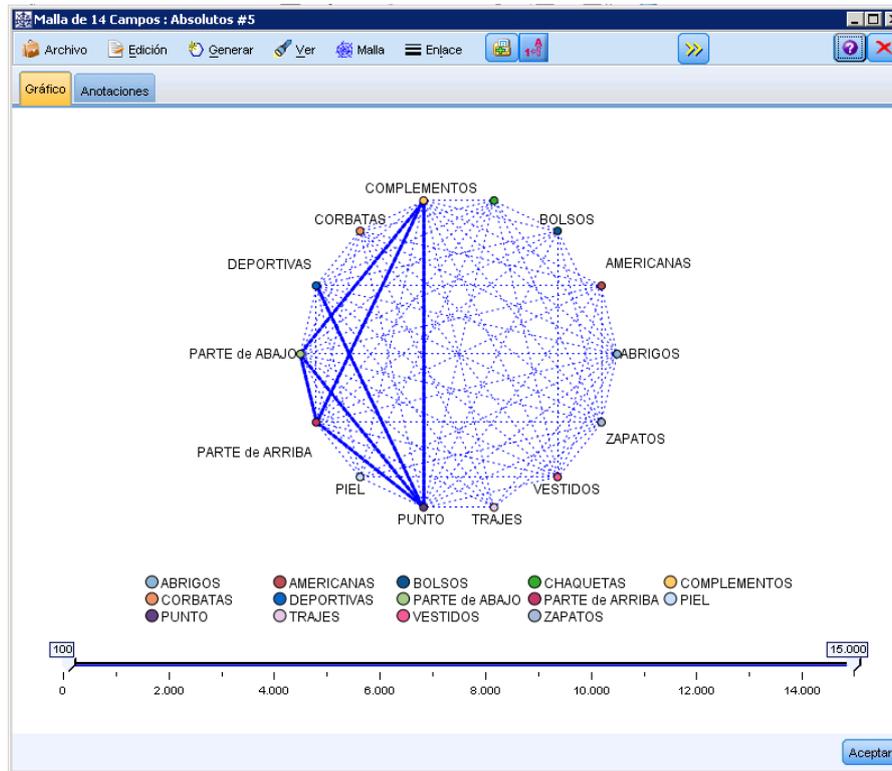
Aplicando el modelo sobre las 14 variables de productos y sobre los 53.382 registros, se obtienen los siguientes resultados:



(Figura 33)

Con una fiabilidad de un 80,1% y aplicable a un 13,3% de los datos, se podría decir que el socio que compra punto, también compra complementos y parte de abajo.

En realidad con este modelo sólo se está identificando un grupo muy definido. Como lo que se está buscando es encontrar más relaciones entre las familias de producto, se aplica sobre los datos un gráfico de malla con el fin de mostrar todas las asociaciones de las familias:



(Figura 34)

Las líneas más remarcadas indican los enlaces más fuertes. En cifras se tendría:

Enlaces	Campo 1	Campo 2
14.575	PUNTO = "1"	PARTE de ABAJO = "1"
14.073	PUNTO = "1"	PARTE de ARRIBA = "1"
10.525	PARTE de ARRIBA = "1"	PARTE de ABAJO = "1"
9.682	PUNTO = "1"	COMPLEMENTOS = "1"
7.087	COMPLEMENTOS = "1"	PARTE de ABAJO = "1"
6.860	COMPLEMENTOS = "1"	PARTE de ARRIBA = "1"
5.198	PUNTO = "1"	DEPORTIVAS = "1"

(Figura 35)

Visualizando el gráfico de malla se puede agrupar a los socios en los siguientes grupos:

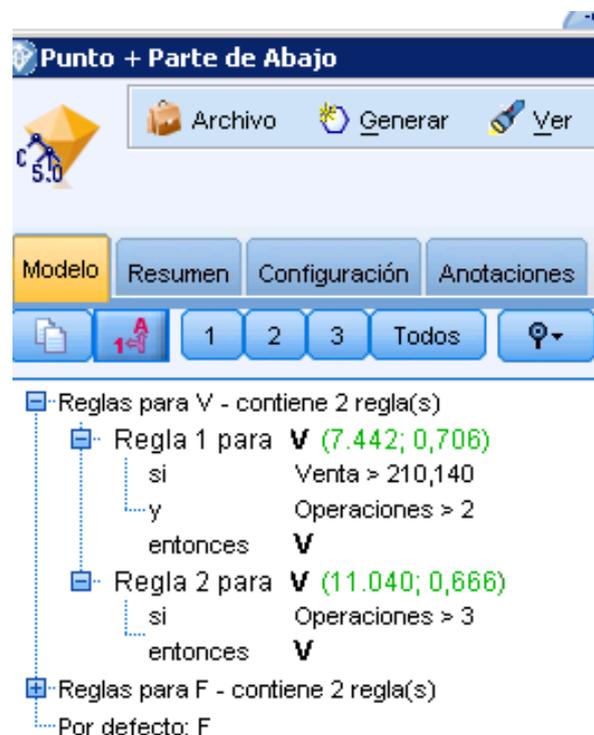
- Compradores Punto + parte de Abajo
- Compradores Punto + parte de Arriba
- Compradores Parte de Arriba + Parte de Abajo
- Compradores Punto + Complementos
- Compradores Complementos + Parte de Abajo
- Compradores Complementos + Parte de Arriba
- Compradores Punto + Deportivas
- Compradores Punto + Parte de Abajo + Parte de Arriba
- Compradores Parte de Arriba + Parte de Abajo + Complementos
- Compradores Parte de Abajo + Complementos + Punto
- Compradores Parte de Arriba + Complementos + Punto
- Compradores Parte de abajo + Parte de arriba + Complementos + Punto

Estos grupos de familias son los grupos que se deberían utilizar en la realización de la campaña.

Para cada uno de estos grupos de socios o segmentos, se podría perfilar aún más la segmentación. Para ello, se busca un conjunto de reglas a través de un árbol de decisión C.5.

Se toma como ejemplo el segmento “Compradores Punto + Parte de Abajo” para realizar el modelo de árbol. Las variables de entrada que se tienen son: sexo, provincia, venta y operaciones.

Los resultados obtenidos son:



(Figura 36)

## 5.4. Segmentación

Actualmente, las promociones son principalmente genéricas y dirigidas sobre el total de la base de datos, donde el incentivo es prácticamente igual para todos los socios. En el área de la segmentación, los resultados de los test promocionales realizados en este ejercicio y anteriores ayudan a actuar sobre el ciclo de vida del socio para la próxima temporada, donde se quiere diferenciar y personalizar mediante propuestas promocionales.

Se propone crear una campaña personalizada al socio en función de su compra, de manera que se le entregará un incentivo para futuras compras realizadas en las secciones para las que se ha visto que dicho socio es comprador habitual. Con esta acción se pretende aumentar el ticket medio y las operaciones del socio, afianzando así su confianza en la marca.

La campaña se realiza en las zonas de Asturias y Murcia. A partir del comportamiento de compra analizado, se tienen 13 grupos de socios diferentes para tener en cuenta en el momento de realizar la segmentación de la campaña. A cada grupo de socios se le enviará un incentivo diferente, y habrá un grupo de socios que no está incluido en las 13 categorías seleccionadas, y recibirán un modelo de campaña genérico.

Estos son los diferentes modelos de la campaña:

1. Promoción habitual, sin incentivo
2. Promoción con incentivo de Complementos
3. Promoción con incentivo de Deportivas
4. Promoción con incentivo de Pantalón/falda
5. Promoción con incentivo de Camisa/blusa
6. Promoción con incentivo de Camisa/blusa + pantalón/falda
7. Promoción con incentivo de Punto
8. Promoción con incentivo de Punto + complementos
9. Promoción con incentivo de Punto + pantalón/falda
10. Promoción con incentivo de Punto + pantalón/falda + complementos
11. Promoción con incentivo de Punto + camisa/blusa
12. Promoción con incentivo de Punto + camisa/blusa + complementos
13. Promoción con incentivo de Punto + camisa/blusa + pantalón/falda
14. Promoción con incentivo de Punto + camisa/blusa + pantalón/falda + complementos

Una vez decidida la segmentación, se procederá a preparar la campaña; se crea a través de un gestor de campañas denominado *Teradata Relationship Manager*, en adelante TRM.

La campaña consta de dos partes: la segmentación y la comunicación. Dentro de la segmentación se realizan las agrupaciones de socios que se desean tener, y en la comunicación se definen los diferentes incentivos que recibirá cada grupo de socios.



(Figura 37)

La campaña se comunica mediante tres canales: postal, e-mail y sms, de manera que se realiza una comunicación de inicio de la campaña y otra de recordatorio del fin de la misma.

El número de socios al que se llegará mediante cada una de las comunicaciones lo marca el presupuesto del que se dispone. En este caso, se tiene:

Inicio Campaña			Fin Campaña	
postal	e-mail	sms	e-mail	sms
57.623	9.227	35.216	8.806	35.216

(Figura 38)

Para crear los segmentos, se necesita de nuevo la ayuda de SQL Assistant debido a que el gestor de campañas no recoge las combinaciones de atributos que en este caso son necesarias. Se lanza una query mediante la cual se genera un listado de socios, que se cargara en la base de datos para poder trabajar con ella en TRM. Para cada segmento se identificará un grupo de control, sin comunicación, que constituye una muestra de la población y debe ser representativo estadísticamente ( $\pm 5\%$ ). Eso permite medir el efecto incremental de la campaña. Además, como se trata de un test, se creará un grupo espejo, que recibe la oferta habitual; este grupo es igualmente una muestra de la población y de la misma manera debe ser representativo ( $\pm 5\%$ ), y va a permitir medir el incremental por oferta.

Para determinar el tamaño muestral, tanto de los grupos de control como de los grupos espejo, se aplica una la fórmula vista en el anexo 1, válida para poblaciones finitas en las cuales N es suficientemente grande ( $>30$ ):

$$n = \frac{Z^2 \cdot N \cdot p \cdot q}{e^2 \cdot (N-1) + Z^2 \cdot p \cdot q}$$

En el cálculo de los grupos espejo, se selecciona un mismo porcentaje para la muestra, del 30%, que en ninguno de los casos supera el error recomendado (2-7%).

El resultado es el siguiente:

Niveles de Confianza		Z	Z^2
IC = 95%		1,96	3,8416

	N	p		N*(Z^2)*P*Q	(Error^2)*(N-1)/(Z^2)*P*Q	n	%	Nº Envíos
		Estimación IDR	% Limite de Error					
Test Solo Complem.	2.031							1.446
Test Solo Complem. (espejo)	2.031	50,0%	3,3%	1950,57240	3,17107	615	30%	615
Test Solo Deportivas	1.225							860
Test Solo Deportivas (espejo)	1.225	50,0%	4,3%	1176,49000	3,22358	365	30%	365
Test Solo Pantalón/Falda	2.925							2.061
Test Solo Pantalón/Falda (espejo)	2.925	50,0%	2,8%	2809,17000	3,25282	864	30%	864
Test Solo Camisa/Blusa	2.637							1.840
Test Solo Camisa/Blusa (espejo)	2.637	50,0%	2,9%	2532,57480	3,17728	797	30%	797
Test Camisa/Blusa + Pantalón/Falda	1.153							806
Test Camisa/Blusa + Pantalón/Falda (espejo)	1.153	50,0%	4,4%	1107,34120	3,19067	347	30%	347
Test Solo Punto	6.979							4.898
Test Solo Punto (espejo)	6.979	50,0%	1,8%	6702,63160	3,22127	2.081	30%	2.081
Test Punto+Complem.	1.683							1.178
Test Punto+Complem. (espejo)	1.683	50,0%	3,7%	1616,35320	3,20124	505	30%	505
Test Punto+Pantalón/Falda	2.811							1.958
Test Punto+Pantalón/Falda (espejo)	2.811	50,0%	2,8%	2699,68440	3,16344	853	30%	853
Test Punto+Pantalón/Falda+Complem	1.522							1.059
Test Punto+Pantalón/Falda+Complem (espejo)	1.522	50,0%	3,8%	1461,72880	3,15672	463	30%	463
Test Punto+Camisa/Blusa	2.605							1.811
Test Punto+Camisa/Blusa (espejo)	2.605	50,0%	2,9%	2501,84200	3,15036	794	30%	794
Test Punto+Camisa/Blusa+Complem	1.234							868
Test Punto+Camisa/Blusa+Complem (espejo)	1.234	50,0%	4,3%	1185,13360	3,24022	366	30%	366
Test Punto+Camisa/Blusa+Pant/Falda	2.407							1.684
Test Punto+Camisa/Blusa+Pant/Falda (espejo)	2.407	50,0%	3,1%	2311,68280	3,19858	723	30%	723
Test Punto+Cam/Blusa+Pant/Falda+Comp	2.272							1.546
Test Punto+Cam/Blusa+Pant/Falda+Comp (espejo)	2.272	50,0%	3,0%	2182,02880	3,00430	726	32%	726
Activos dec 1-5	10.351	50,0%	4,0%	9941,10040	17,52040	567	5%	9.784
Activos dec 6-10	8.349	50,0%	4,5%	8018,37960	17,86510	449	5%	7.900
Inact. Recientes	70	50,0%	5,0%	67,22800	1,13290		0%	70
Aband. Recientes	8.835	50,0%	4,5%	8485,13400	18,84925	450	5%	8.385

Es el grupo de control calculado para cada segmento, válido para analizar tanto el segmento de test, como el segmento espejo.

Seleccionamos el mismo número de socios para grupo espejo.

El límite de error en ninguno de los casos supera el error recomendado.

(Figura 39)

Estas cifras se introducen en el gestor de campañas y a continuación se lanza la campaña con la posterior extracción de los ficheros. Se obtienen 14 ficheros distintos, 13 que se corresponden con las categorías de productos seleccionadas anteriormente, y uno con la promoción genérica, que se envía al resto de socios y a los grupos espejo.

## 5.5. Análisis de resultados

El análisis de la campaña se realiza mediante la ejecución de un informe con Microstrategy, manipulado en un fichero Excel. Consiste, por un lado, en la obtención de resultados generales de la campaña, en cuanto a determinadas métricas calculadas, y por otro lado, en el estudio de la rentabilidad de la comunicación de la campaña.

Definición de las métricas que se calcularán:

Envíos	número de socios que ha recibido la comunicación.
Compradores	número de socios comunicados que han comprado durante la promoción.
IR	porcentaje de socios comunicados que han comprado durante la promoción (Compradores/Envíos).
Dcto	valor de descuento que se ha aplicado a los socios compradores.
VN	valor de las compras realizadas por los compradores durante la promoción.
Contrib.	valor neto obtenido por las compras.
VNxCompr.	media de compra realizada por los socios (VN/Compradores).
Contrib.xCompr.	media del valor neto obtenido por comprador (Contrib./Compradores).
Op.	número de operaciones realizadas por los socios.
Frec.	frecuencia de compra por comprador (Op./Compradores).
Ud.	unidades vendidas.
Art.xOp.	media de artículos vendidos por operación (Ud./Op.).
Costes envío	costes que conlleva el envío de las comunicaciones realizadas a lo largo de la campaña.
ROI	retorno sobre la inversión (Contrib./Costes Envío).

(Figura 40)

De esta manera, obtenemos que los resultados generales muestran que con la campaña de producto se obtiene un índice de respuesta superior al obtenido con la campaña habitual, es decir, el efecto de la comunicación ha sido mayor en el caso de incluir el incentivo de producto:

Análisis Promoción ASMU																
Fecha análisis: Entre 07/11/2013 y 17/11/2013																
Fuente: Dpto. Customer Intelligence																
CTL_IND	N															
Campaña	(Todas)															
1ª Comunicación	S															
Agr.Promo	Envíos	Compra dores	IR%	Dcto.	% Dcto.	V.N.	Contrib.	% MB	VN. x Compr.	Contrib. x Compr.	Op.	Frec.	Ud.	Art. x Op.	costes envío	ROI
Producto	21.985	1.601	7,3%	43.023 €	25,1%	128.099 €	73.245 €	57,2%	80,0 €	45,7 €	1.842	1,2	3.238	1,76	6.156 €	10,9 €
Dcto. Incremental	35.634	1.919	5,4%	52.449 €	25,3%	154.492 €	89.116 €	57,7%	80,5 €	46,4 €	2.199	1,1	3.964	1,80	9.978 €	7,9 €
<b>Total general</b>	<b>57.619</b>	<b>3.520</b>	<b>6,1%</b>	<b>95.472 €</b>	<b>25,3%</b>	<b>282.591 €</b>	<b>162.361 €</b>	<b>57,5%</b>	<b>80,3 €</b>	<b>46,1 €</b>	<b>4.041</b>	<b>1,1</b>	<b>7.202</b>	<b>1,78</b>	<b>16.133 €</b>	<b>9,1 €</b>

(Figura 41)

Si se desglosan estos resultados para ver de manera más detallada la segmentación, se obtienen unos índices de respuesta bastante favorables en los grupos de: Punto + Pantalón/Falda + Complementos, Punto + Camisa/Blusa + Pantalón/Falda y Punto + Camisa/Blusa + Pantalón/Falda + Complementos:

Análisis Promoción ASMU																
Fecha análisis: Entre 07/11/2013 y 17/11/2013																
Fuente: Dpto. Customer Intelligence																
CTL_IND	N															
Campaña	(Todas)															
1ª Comunicación	S															
modelo	Envíos	Compras	IR%	Dcto.	% Dcto.	V.N.	Contrib.	% MB	VN. x Compr.	Contrib. x Compr.	Op.	Frec.	Ud.	Art. x Op.	costes envío	ROI
Sólo complementos	1.416	51	3,6%	1.285 €	23,9%	4.091 €	2.151 €	52,6%	80,2 €	42,2 €	55	1,1	98	1,78	396 €	4,4 €
Sólo deportivas	860	26	3,0%	600 €	25,3%	1.773 €	1.071 €	60,4%	68,2 €	41,2 €	33	1,3	51	1,55	241 €	3,4 €
Sólo Pantalón/Falda	2.061	101	4,9%	2.066 €	22,5%	7.096 €	4.179 €	58,9%	70,3 €	41,4 €	111	1,1	175	1,58	577 €	6,2 €
Sólo Camisa/Blusa	1.840	73	4,0%	1.957 €	25,4%	5.754 €	3.318 €	57,7%	78,8 €	45,5 €	83	1,1	136	1,64	515 €	5,4 €
Camisa/Blusa + Pantalón/Falda	806	53	6,6%	1.404 €	24,8%	4.265 €	2.440 €	57,2%	80,5 €	46,0 €	60	1,1	99	1,65	226 €	9,8 €
Sólo Punto	4.898	228	4,7%	4.544 €	23,8%	14.553 €	8.412 €	57,8%	63,8 €	36,9 €	254	1,1	404	1,59	1.371 €	5,1 €
Punto + Complementos	1.178	78	6,6%	1.940 €	24,6%	5.945 €	3.626 €	61,0%	76,2 €	46,5 €	87	1,1	129	1,48	330 €	10,0 €
Punto + Pantalón/Falda	1.958	155	7,9%	4.192 €	24,7%	12.798 €	7.021 €	54,9%	82,6 €	45,3 €	175	1,1	309	1,77	548 €	11,8 €
Punto + Pantalón/Falda + Complementos	1.059	124	11,7%	2.680 €	21,9%	9.558 €	5.515 €	57,7%	77,1 €	44,5 €	140	1,1	242	1,73	297 €	17,6 €
Punto + Camisa/Blusa	1.811	126	7,0%	3.097 €	24,9%	9.352 €	5.624 €	60,1%	74,2 €	44,6 €	141	1,1	241	1,71	507 €	10,1 €
Punto + Camisa/Blusa + Complementos	868	86	9,9%	2.516 €	24,4%	7.808 €	4.499 €	57,6%	90,8 €	52,3 €	103	1,2	202	1,96	243 €	17,5 €
Punto + Camisa/Blusa + Pantalón/Falda	1.684	194	11,5%	4.766 €	25,3%	14.106 €	8.140 €	57,7%	72,7 €	42,0 €	227	1,2	406	1,79	472 €	16,3 €
Punto + Camisa/Blusa + Pantalón/Falda + Complementos	1.546	306	19,8%	11.977 €	27,9%	30.999 €	17.249 €	55,6%	101,3 €	56,4 €	373	1,2	746	2,00	433 €	38,8 €
Dcto. Incremental General	35.634	1.919	5,4%	52.449 €	25,3%	154.492 €	89.116 €	57,7%	80,5 €	46,4 €	2.199	1,1	3.964	1,80	9.978 €	7,9 €
<b>Total general</b>	<b>57.619</b>	<b>3.520</b>	<b>6,1%</b>	<b>95.472 €</b>	<b>25,3%</b>	<b>282.591 €</b>	<b>162.361 €</b>	<b>57,5%</b>	<b>80,3 €</b>	<b>46,1 €</b>	<b>4.041</b>	<b>1,1</b>	<b>7.202</b>	<b>1,78</b>	<b>16.133 €</b>	<b>9,1 €</b>

(Figura 42)

A nivel general, se compara también el resultado de la promoción en las regiones de test y en el resto de España, donde solamente se envió la comunicación genérica. Viendo la respuesta se puede deducir que el índice de respuesta de los socios en ambas zonas fue muy bueno, siendo superior en las regiones de test; también es superior la venta y contribución por comprador:

Canal de comunicación: E-mail (E), postal (P) y sms (S).

AsturiasMurcia	colateral	Envíos	Compradores	IR%	Dcto.	% Dcto.	V.N.	Contrib.	% MB	VN. x Compr.	Contrib. x Compr.	Op.	Frec.	Ud.	Art. x Op.	costes envío	ROI
Asturias y Murcia	E	21.246	464	2,2%	10.723 €	27,8%	27.842 €	15.965 €	57,3%	60,0 €	34,4 €	521	1,1	788	1,51	21 €	750,4 €
	P	33.667	3.206	9,5%	97.927 €	29,2%	236.943 €	135.827 €	57,3%	73,9 €	42,4 €	3.991	1,2	6.439	1,61	14.073 €	8,7 €
	S	32.508	1.073	3,3%	23.125 €	27,4%	61.300 €	35.689 €	58,2%	57,1 €	33,3 €	1.210	1,1	1.738	1,44	1.463 €	23,4 €
<b>Total Asturias Murcia</b>		<b>87.421</b>	<b>4.743</b>	<b>5,4%</b>	<b>131.774 €</b>	<b>28,8%</b>	<b>326.084 €</b>	<b>187.482 €</b>	<b>57,5%</b>	<b>68,8 €</b>	<b>39,5 €</b>	<b>5.722</b>	<b>1,2</b>	<b>8.965</b>	<b>1,57</b>	<b>15.557 €</b>	<b>11,1 €</b>
Resto	E	353.189	7.136	2,0%	172.439 €	27,8%	446.905 €	254.703 €	57,0%	62,6 €	35,7 €	8.077	1,1	12.885	1,60	353 €	720,2 €
	P	467.172	38.491	8,2%	1.148.580 €	29,8%	2.710.870 €	1.547.103 €	57,1%	70,4 €	40,2 €	47.213	1,2	77.120	1,63	195.278 €	6,9 €
	S	375.021	12.257	3,3%	275.655 €	27,8%	716.799 €	416.663 €	58,1%	58,5 €	34,0 €	13.762	1,1	20.989	1,52	16.876 €	23,7 €
<b>Total Resto</b>		<b>1.195.382</b>	<b>57.884</b>	<b>4,8%</b>	<b>1.696.654 €</b>	<b>29,2%</b>	<b>3.874.574 €</b>	<b>2.218.469 €</b>	<b>57,3%</b>	<b>66,9 €</b>	<b>38,3 €</b>	<b>69.052</b>	<b>1,2</b>	<b>110.974</b>	<b>1,61</b>	<b>212.507 €</b>	<b>9,4 €</b>
<b>Total general</b>		<b>1.282.803</b>	<b>62.627</b>	<b>4,9%</b>	<b>1.728.428 €</b>	<b>29,2%</b>	<b>4.200.658 €</b>	<b>2.405.950 €</b>	<b>57,3%</b>	<b>67,1 €</b>	<b>38,4 €</b>	<b>74.774</b>	<b>1,2</b>	<b>119.939</b>	<b>1,60</b>	<b>228.064 €</b>	<b>9,5 €</b>

(Figura 43)

En cuanto a la rentabilidad de la campaña por canal de comunicación, se considera como el apartado más importante del análisis; consiste en comparar el comportamiento de los socios comunicados y los no comunicados (reservados como grupo de control), estimar qué hubiera ocurrido si la campaña no se hubiese enviado y ver el efecto incremental obtenido debido al envío de la comunicación.

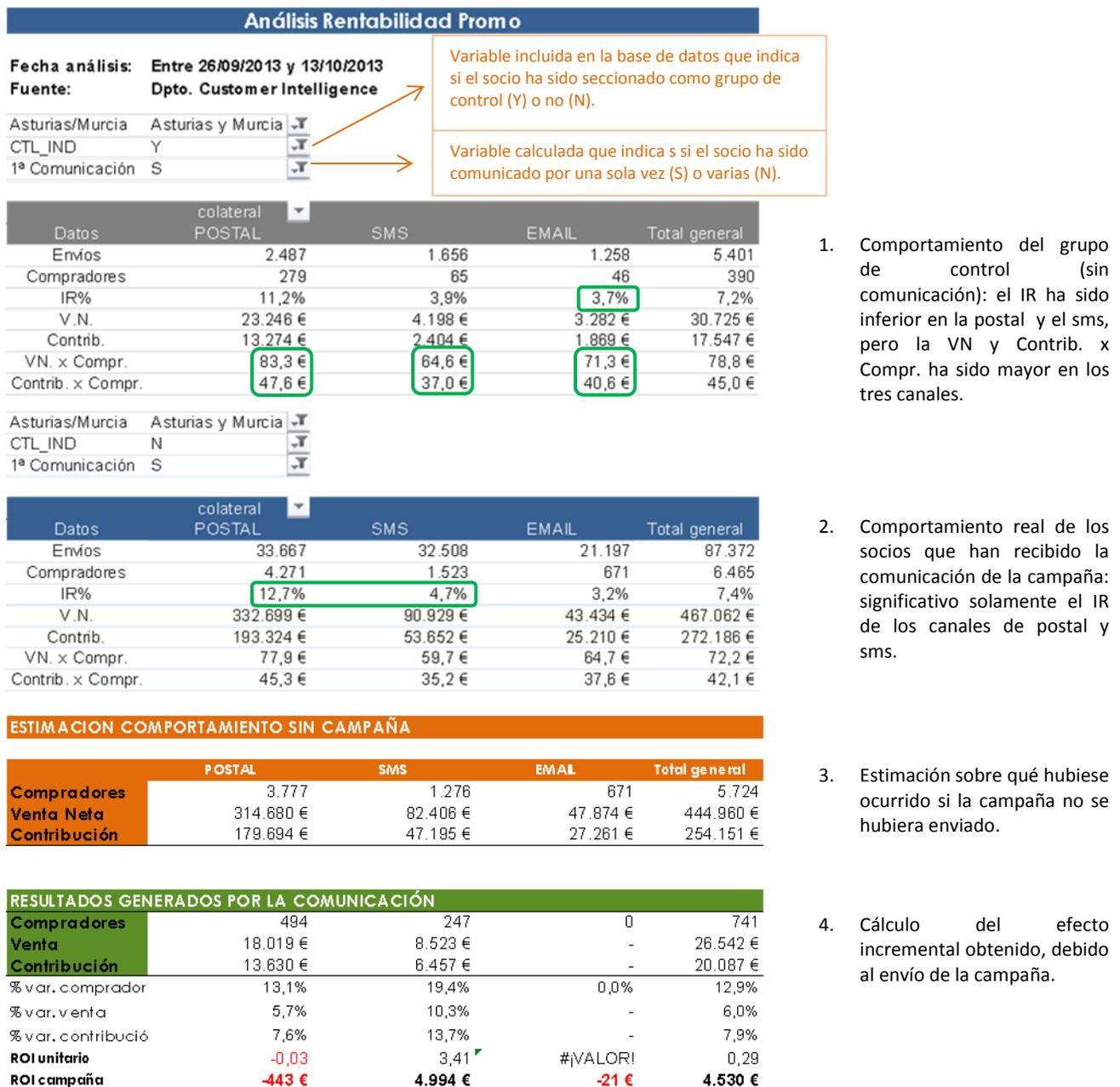
Se va a ver la rentabilidad de la postal, del e-mail y del sms de dos maneras. En primer lugar analizará cada canal de manera conjunta, y después se verá por canal de comunicación cada grupo de socios o segmento.

En el análisis de la rentabilidad por canal de comunicación se comparará cada segmento con su grupo de control asociado, para ver si la campaña ha sido efectiva, y por otro lado, se compararán dichos grupos de socios con su grupo espejo correspondiente, para comprobar si los test realizados han sido satisfactorios.

## Rentabilidad por canal de socios comunicados frente a su grupo de control:

Se analiza la rentabilidad de cada uno de los canales, comparando el comportamiento de todos los socios a los que se comunicó la campaña con los socios que reservaron como grupo de control, a los que no se les envió ningún tipo de comunicación.

Se muestra el significado de cada uno de los apartados del análisis:



(Figura 44)

El resultado es positivo para canal de comunicación sms, pero no para la postal y el e-mail.

## Rentabilidad de producto:

Se estudia la rentabilidad de la comunicación con incentivo, de esta manera se muestra comportamiento de los socios a los que se envió la comunicación con el incentivo, frente a sus grupos espejo a los que se les envió la comunicación genérica, sin incentivo. Se diferencia el canal de comunicación.

### POSTAL:

Fecha análisis: Entre 07/11/2013 y 17/11/2013  
Fuente: Dpto. Customer Intelligence

Grupo Espejo	Espejo	SEG_NM													
colateral	P	Solo Complem. (espejo)													
		Solo Deportivas (espejo)	Solo Pantalón/Falda (espejo)	Solo Camisa/Blusa (espejo)	Camisa/Blusa + Pantalón/Falda (espejo)	Solo Punto (espejo)	Punto+Complem. (espejo)	Punto+Pantalón/Falda (espejo)	Punto+Pantalón/Falda+Complem. (espejo)	Punto+Camisa/Blusa (espejo)	Punto+Camisa/Blusa+Complem. (espejo)	Punto+Camisa/Blusa+Pant./Falda (espejo)	Punto+Cam/Blusa+Pant./Falda+Comp. (espejo)	Total general	
Datos		615	365	864	797	347	2.080	505	853	463	793	366	723	726	9.497
Emisos		35	12	33	30	24	98	36	64	57	38	74	123	123	662
IR%		5,7%	3,3%	3,8%	3,8%	6,9%	4,7%	7,1%	7,5%	12,3%	4,8%	10,4%	10,2%	16,9%	7,0%
V.N.		2.187 €	1.697 €	1.865 €	2.630 €	2.001 €	7.578 €	2.948 €	3.853 €	2.001 €	3.186 €	5.879 €	10.158 €	50.241 €	50.241 €
Contrib.		1.233 €	929 €	1.126 €	1.579 €	1.216 €	4.458 €	1.591 €	2.487 €	2.252 €	1.988 €	3.364 €	5.761 €	29.116 €	29.116 €
VN. x Compr.		62,5 €	141,4 €	56,5 €	87,7 €	85,4 €	77,3 €	81,9 €	66,2 €	67,6 €	53,2 €	83,8 €	79,4 €	82,6 €	75,9 €
Contrib. x Compr.		35,2 €	77,4 €	34,1 €	52,8 €	50,7 €	45,5 €	44,2 €	38,9 €	39,5 €	52,3 €	45,5 €	46,8 €	44,0 €	44,0 €
Costes envío		172 €	102 €	242 €	223 €	97 €	582 €	141 €	239 €	130 €	222 €	102 €	202 €	203 €	2.659 €

### ESTIMACION COMPORTAMIENTO SIN CAMPAÑA

Grupo Espejo	Promo	SEG_NM													
colateral	P	Solo Complem.													
		Solo Deportivas	Solo Pantalón/Falda	Solo Camisa/Blusa	Camisa/Blusa + Pantalón/Falda	Solo Punto	Punto+Complem.	Punto+Pantalón/Falda	Punto+Pantalón/Falda+Complem.	Punto+Camisa/Blusa	Punto+Camisa/Blusa+Complem.	Punto+Camisa/Blusa+Pant./Falda	Punto+Cam/Blusa+Pant./Falda+Comp.	Total general	
Datos		1.416	860	2.061	1.840	606	4.898	1.178	1.959	1.059	1.811	868	1.684	1.546	21.985
Emisos		51	26	101	73	53	228	78	155	124	126	86	194	306	1.601
IR%		3,6%	3,0%	4,9%	4,0%	6,6%	4,7%	6,6%	7,9%	11,7%	7,0%	9,9%	11,5%	19,8%	7,3%
V.N.		4.091 €	1.773 €	7.096 €	5.754 €	4.265 €	14.553 €	5.945 €	12.798 €	9.558 €	9.352 €	7.808 €	14.106 €	30.999 €	128.099 €
Contrib.		2.151 €	1.071 €	4.179 €	3.318 €	2.440 €	8.412 €	3.626 €	7.021 €	5.515 €	5.624 €	4.499 €	8.140 €	17.249 €	73.245 €
VN. x Compr.		88,2 €	70,3 €	78,8 €	80,3 €	83,8 €	79,2 €	82,6 €	77,1 €	74,2 €	90,8 €	72,7 €	101,3 €	80,0 €	80,0 €
Contrib. x Compr.		42,2 €	41,4 €	45,2 €	46,5 €	46,0 €	36,9 €	46,5 €	45,3 €	44,5 €	44,6 €	52,3 €	42,0 €	45,4 €	45,7 €
Costes envío		386 €	241 €	577 €	515 €	226 €	1.371 €	330 €	548 €	297 €	507 €	243 €	472 €	433 €	6.156 €

### RESULTADOS GENERADOS POR LA COMUNICACIÓN

Compradores	51	26	79	69	53	228	78	147	124	87	86	172	262	1.462
Venta Neta	3.186 €	3.678 €	4.448 €	6.073 €	4.419 €	17.630 €	6.388 €	9.731 €	8.382 €	4.614 €	7.209 €	13.893 €	21.631 €	111.082 €
Contribución	1.797 €	2.013 €	2.685 €	3.646 €	2.685 €	10.371 €	3.447 €	5.710 €	4.898 €	2.587 €	4.498 €	7.835 €	12.268 €	64.441 €

Compradores	0	0	22	4	0	0	0	8	0	39	0	22	44	139
Venta	-	-	2.648 €	-318 €	-	-	-	3.067 €	-	4.738 €	-	413 €	9.368 €	19.915 €
Contribución	-	-	1.494 €	-328 €	-	-	-	1.311 €	-	3.037 €	-	304 €	4.982 €	10.800 €
% var. compradore	0,0%	0,0%	28,3%	5,4%	0,0%	0,0%	0,0%	5,5%	0,0%	45,2%	0,0%	12,6%	16,8%	9,5%
% var. venta	-	-	59,5%	-5,2%	-	-	-	31,5%	-	102,7%	-	3,0%	43,3%	17,9%
% var. contribución	-	-	55,6%	-9,0%	-	-	-	23,0%	-	117,4%	-	3,9%	40,6%	16,8%
ROI unitario	# VALOR!	# VALOR!	1,59	-1,64	# VALOR!	# VALOR!	# VALOR!	1,39	# VALOR!	4,99	# VALOR!	-0,35	10,51	0,75
ROI campaña	-398 €	-241 €	917 €	-843 €	-226 €	-1.371 €	-330 €	763 €	-297 €	2.530 €	-243 €	-167 €	4.549 €	4.644 €

Result. Campaña	No significativo	No significativo	Resultado Positivo	No significativo	No significativo	No significativo	No significativo	Resultado Positivo	No significativo	Resultado Positivo	No significativo	No significativo	Resultado Positivo	Resultado Positivo
-----------------	------------------	------------------	--------------------	------------------	------------------	------------------	------------------	--------------------	------------------	--------------------	------------------	------------------	--------------------	--------------------

(Figura 45)

### SMS:

Fecha análisis: Entre 07/11/2013 y 17/11/2013  
Fuente: Dpto. Customer Intelligence

Grupo Espejo	Espejo	SEG_NM													
colateral	S	Comunic. Solo Complem.													
		Comunic. Solo Deportivas	Comunic. Solo Pantalón/Falda	Comunic. Solo Camisa/Blusa	Comunic. Camisa/Blusa + Pantalón/Falda	Comunic. Solo Punto	Comunic. Punto+Complem.	Comunic. Punto+Pantalón/Falda	Comunic. Punto+Pantalón/Falda+Complem.	Comunic. Punto+Camisa/Blusa	Comunic. Punto+Camisa/Blusa+Complem.	Comunic. Punto+Camisa/Blusa+Pant./Falda	Comunic. Punto+Cam/Blusa+Pant./Falda+Comp.	Total general	
Datos		213	114	277	263	110	679	160	295	167	272	140	257	243	3.190
Emisos		10	3	7	6	4	19	8	12	10	6	13	24	129	129
IR%		4,7%	2,6%	2,5%	2,3%	3,6%	2,8%	5,0%	4,1%	6,0%	2,2%	5,1%	9,8%	4,0%	4,0%
V.N.		849 €	370 €	573 €	748 €	370 €	1.263 €	1.084 €	967 €	603 €	192 €	508 €	1.468 €	2.207 €	11.289 €
Contrib.		560 €	188 €	368 €	451 €	208 €	773 €	458 €	546 €	384 €	327 €	806 €	1.229 €	6.389 €	6.389 €
VN. x Compr.		54,9 €	123,2 €	81,8 €	124,6 €	92,5 €	88,5 €	135,5 €	79,7 €	60,3 €	32,0 €	72,6 €	112,9 €	91,9 €	87,5 €
Contrib. x Compr.		55,0 €	62,6 €	51,1 €	75,1 €	52,1 €	40,7 €	57,2 €	45,5 €	38,4 €	15,1 €	46,7 €	62,0 €	51,2 €	49,4 €
Costes envío		10 €	5 €	12 €	12 €	5 €	31 €	7 €	13 €	8 €	12 €	6 €	12 €	11 €	144 €

Grupo Espejo	Promo	SEG_NM													
colateral	S	Comunic. Solo Complem.													
		Comunic. Solo Deportivas	Comunic. Solo Pantalón/Falda	Comunic. Solo Camisa/Blusa	Comunic. Camisa/Blusa + Pantalón/Falda	Comunic. Solo Punto	Comunic. Punto+Complem.	Comunic. Punto+Pantalón/Falda	Comunic. Punto+Pantalón/Falda+Complem.	Comunic. Punto+Camisa/Blusa	Comunic. Punto+Camisa/Blusa+Complem.	Comunic. Punto+Camisa/Blusa+Pant./Falda	Comunic. Punto+Cam/Blusa+Pant./Falda+Comp.	Total general	
Datos		444	285	596	623	271	1.579	389	673	367	629	331	586	544	7.305
Emisos		14	4	15	15	6	49	12	27	28	19	15	36	63	303
IR%		3,2%	1,4%	2,5%	2,4%	2,2%	3,1%	3,1%	4,0%	7,6%	3,0%	4,5%	6,1%	11,6%	4,1%
V.N.		1.887 €	203 €	1.042 €	1.505 €	429 €	3.950 €	2.411 €	2.642 €	1.195 €	2.862 €	1.285 €	2.693 €	6.506 €	28.206 €
Contrib.		920 €	198 €	616 €	837 €	271 €	2.171 €	685 €	441 €	1.289 €	1.620 €	732 €	1.490 €	3.547 €	14.784 €
VN. x Compr.		133,4 €	73,3 €	69,4 €	100,3 €	71,6 €	74,5 €	62,8 €	88,9 €	84,4 €	62,9 €	85,7 €	73,3 €	103,3 €	86,5 €
Contrib. x Compr.		65,7 €	49,0 €	41,2 €	58,8 €	45,1 €	44,3 €	36,7 €	44,7 €	57,8 €	38,5 €	49,7 €	74,4 €	56,3 €	48,8 €
Costes envío		20 €	15 €	27 €	28 €	12 €	71 €	17 €	30 €	16 €	28 €	15 €	28 €	24 €	329 €

### ESTIMACION COMPORTAMIENTO SIN CAMPAÑA

Compradores	14	4	15	14	6	44	12	27	21	14	15	30	54	270
Venta Neta	1.329 €	493 €	1.227 €	1.771 €	555 €	2.936 €	1.626 €	2.153 €	1.288 €	444 €	1.089 €	3.346 €	4.940 €	23.197 €
Contribución	771 €	250 €	787 €	1.088 €	313 €	1.788 €	687 €	1.228 €	822 €	209 €	701 €	1.838 €	2.750 €	13.201 €

Compradores	0	0	0	1	0	5	0	0	7	5	0	6	9	33
Venta	-	-	-	-298 €	-	713 €	-	-	1.354 €	751 €	-	704 €	1.566 €	3.409 €
Contribución	-	-	-	-231 €	-	373 €	-	-	788 €	623 €	-	-348 €	798 €	1.911 €
% var. compradore	0,0%	0,0%	0,0%	5,5%	0,0%	10,9%	0,0%	0,0%	31,0%	36,9%	0,0%	21,4%	17,3%	12,2%
% var. venta	-	-	-	-15,0%	-	24,3%	-	-	105,1%	168,3%	-	-21,2%	31,7%	14,7%
% var. contribución	-	-	-	-21,6%	-	20,7%	-	-	97,2%	250,1%	-	-18,9%	29,0%	14,5%
ROI unitario	# VALOR!	# VALOR!	# VALOR!	-0,25	# VALOR!	4,25	# VALOR!	# VALOR!	48,70	17,48	# VALOR!	-14,21	31,53	4,81
ROI campaña	-20 €	-13 €	-27 €	-289 €	-12 €	302 €	-17 €	-36 €	782 €	495 €	-16 €	-375 €	772 €	1.882 €

Result. Campaña	No significativo	Resultado Positivo	No significativo	No significativo	Resultado Positivo	Resultado Positivo	No significativo	No significativo	Resultado Positivo	Resultado Positivo				
-----------------	------------------	------------------	------------------	------------------	------------------	--------------------	------------------	------------------	--------------------	--------------------	------------------	------------------	--------------------	--------------------

EMAIL:

Fecha análisis: Entre 07/11/2013 y 17/11/2013  
Fuente: Depto. Customer Intelligence

Grupo Espejo	Espejo	<input checked="" type="checkbox"/>
colateral	E	<input checked="" type="checkbox"/>

Datos	Comunic. Solo Complet. (espejo)	Comunic. Solo Deportivas (espejo)	Comunic. Solo Pantalón/Falda (espejo)	Comunic. Solo Camisa/Blusa (espejo)	Comunic. Camis/Blusa+Pant /Falda (espejo)	Comunic. Solo Punto (espejo)	Comunic. Punto+Complem. (espejo)	Comunic. Punto+Pantalón/Falda (espejo)	Comunic. Punto+Pant/Falda +Comp (espejo)	Comunic. Punto+Camisa/Blusa (espejo)	Comunic. Punto+Cam/Blusa +Comp (espejo)	Comunic. Pto+Cam/Blus+Pant/Fald (espejo)	Comunic. Pto+CamBlus+Pant/Comp (espejo)	Total general
Envíos	60	27	76	74	44	194	73	101	66	110	50	94	102	1.071
Compradores	2	1	1	1	1	3	2	3	6	1	5	5	8	38
IR%	3,3%	0,0%	1,3%	1,4%	2,3%	1,5%	2,7%	3,0%	9,1%	0,9%	10,0%	5,3%	7,8%	3,5%
V.N.	208 €		66 €	130 €	31 €	270 €	205 €	158 €	155 €	95 €	363 €	305 €	877 €	2.863 €
Contrib.	126 €		37 €	71 €	22 €	135 €	135 €	77 €	87 €	49 €	229 €	166 €	539 €	1.674 €
VN. x Compr.	103,9 €		66,1 €	130,2 €	31,4 €	90,0 €	102,5 €	52,6 €	25,9 €	95,0 €	72,5 €	61,0 €	109,6 €	75,3 €
Contrib. x Compr.	63,2 €		37,2 €	71,4 €	22,4 €	45,2 €	67,4 €	25,8 €	14,6 €	48,8 €	45,9 €	33,2 €	67,3 €	44,1 €
Costes envío	0 €	0 €	0 €	0 €	0 €	0 €	0 €	0 €	0 €	0 €	0 €	0 €	0 €	1 €

Grupo Espejo	Promo	<input checked="" type="checkbox"/>
colateral	E	<input checked="" type="checkbox"/>
Agr.Promo	Producto	<input checked="" type="checkbox"/>

Datos	Comunic. Solo	Total general												
Envíos	145	69	188	171	82	474	146	208	132	225	121	218	244	2.423
Compradores	2	1	5	2	1	14	7	10	7	6	7	13	22	97
IR%	1,4%	1,4%	2,7%	1,2%	1,2%	3,0%	4,8%	4,8%	5,3%	2,7%	5,8%	6,0%	9,0%	4,0%
V.N.	120 €	83 €	537 €	121 €	41 €	1.002 €	479 €	331 €	293 €	472 €	597 €	887 €	2.012 €	6.974 €
Contrib.	74 €	48 €	305 €	53 €	17 €	529 €	258 €	165 €	191 €	309 €	383 €	505 €	1.182 €	4.018 €
VN. x Compr.	59,9 €	82,6 €	107,4 €	60,3 €	41,3 €	71,6 €	68,4 €	33,1 €	41,8 €	78,7 €	85,3 €	68,2 €	91,5 €	71,9 €
Contrib. x Compr.	36,8 €	47,7 €	61,1 €	26,6 €	16,9 €	37,8 €	36,8 €	16,5 €	27,3 €	51,5 €	54,6 €	38,9 €	53,7 €	41,4 €
Costes envío	0 €	0 €	0 €	0 €	0 €	0 €	0 €	0 €	0 €	0 €	0 €	0 €	0 €	2 €

ESTIMACION COMPORTAMIENTO SIN CAMPAÑA

	Comunic. Solo Complet.	Comunic. Solo Deportivas	Comunic. Solo Pantalón/Falda	Comunic. Solo Camisa/Blusa	Comunic. Camis/Blusa + Pantalón/Falda	Comunic. Solo Punto	Comunic. Punto+Complem.	Comunic. Punto+Pantalón/Falda	Comunic. Punto+Pant/Falda+Complem.	Comunic. Punto+Camisa/Blusa	Comunic. Punto+Cam/Blusa +Complem.	Comunic. Punto+Cam/Blusa +Pant/Falda	Comunic. Punto+CamBlus+Pant/Falda+Comp.	Total general
Compradores	2	0	2	2	1	7	4	6	7	2	7	12	19	72
Venta Neta	208 €	0 €	164 €	260 €	31 €	659 €	410 €	325 €	181 €	194 €	508 €	707 €	2.097 €	5.745 €
Contribución	126 €	0 €	92 €	143 €	22 €	331 €	270 €	158 €	102 €	100 €	321 €	384 €	1.288 €	3.337 €

RESULTADOS GENERADOS POR LA COMUNICACIÓN

Compradores	0	1	3	0	0	7	3	4	0	4	0	1	3	25
Venta	-	83 €	374 €	-	-	342 €	69 €	6 €	-	278 €	-	179 €	-85 €	1.246 €
Contribución	-	48 €	213 €	-	-	198 €	-12 €	7 €	-	209 €	-	121 €	-106 €	678 €
% var. compradore	0,0%	-	102,1%	0,0%	0,0%	91,0%	75,0%	61,9%	0,0%	193,3%	0,0%	12,1%	15,0%	35,2%
% var. venta	-	-	228,4%	-	-	51,9%	16,9%	1,9%	-	142,8%	-	25,3%	-4,0%	21,7%
% var. contribución	-	-	232,0%	-	-	60,0%	-4,4%	4,3%	-	209,7%	-	31,4%	-8,2%	20,3%
ROI unitario	# VALOR!	1,078,48	1,772,59	# VALOR!	# VALOR!	653,21	-129,32	49,80	# VALOR!	1,451,46	# VALOR!	863,67	-679,38	436,32
ROI campaña	0 €	48 €	213 €	0 €	0 €	198 €	-12 €	7 €	0 €	209 €	0 €	120 €	-106 €	677 €

Result. Campaña	No significativo	Resultado Positivo	Resultado Positivo	No significativo	No significativo	Resultado Positivo								
-----------------	------------------	--------------------	--------------------	------------------	------------------	--------------------	------------------	--------------------	------------------	--------------------	------------------	--------------------	------------------	--------------------

(Figura 47)

Se obtiene que para los tres canales de comunicación, los resultados son positivos, pero básicamente, los grupos de producto que han reportado incremento de venta son Punto, y las combinaciones de Punto, Pantalón y Falda, Camisas y Blusas y complementos.

## 6. CONCLUSIONES

- Se ha podido determinar que la selección de Asturias y Murcia como zonas de test, no fue acertada porque se ha demostrado que no representa suficientemente las frecuencias de toda España. Lo idóneo sería determinar la muestra representativa mediante un muestreo aleatorio estratificado, en el que los tamaños muestrales de los estratos sean proporcionales a las frecuencias de cada uno de ellos.
- La selección de familias debería ajustarse a:
  - Compradores Punto + parte de Abajo
  - Compradores Punto + parte de Arriba
  - Compradores Parte de Arriba + Parte de Abajo
  - Compradores Punto + Complementos
  - Compradores Complementos + Parte de Abajo
  - Compradores Complementos + Parte de Arriba
  - Compradores Punto + Deportivas
  - Compradores Punto + Parte de Abajo + Parte de Arriba
  - Compradores Parte de Arriba + Parte de Abajo + Complementos
  - Compradores Parte de Abajo + Complementos + Punto
  - Compradores Parte de Arriba + Complementos + Punto
  - Compradores Parte de abajo + Parte de arriba + Complementos + Punto
- La campaña en general ha generado beneficios, especialmente han respondido de manera positiva los socios comunicados mediante canal de comunicación sms.
- Viendo por separado los resultados de la campaña de producto más incentivo, con respecto a su grupo espejo que recibe la campaña genérica, es positiva para algunas de las familias seleccionadas, en concreto el resultado es óptimo para las combinaciones de las secciones de Punto, Camisas y Blusas, Pantalón y Falda y Complementos.
- Como conclusión final, se diría que lo correcto sería elegir una zona de test que sí fuera representativa de la población, y posteriormente crear la campaña de producto en base a los resultados obtenidos en el modelo A Priori y posteriores árboles de decisión.

## 7. BIBLIOGRAFÍA

### Libros:

- Técnicas de Muestreo. William G. Cochran. 1996
- Estadística para Administradores. Richard I. Levin & David S. Rubin. Sexta Edición.
- MicroStrategy Desktop: Diseño Avanzado de Informes. 2005, MicroStrategy, Inc.
- Predictive Modeling with Clementine. 2008 by SPSS Inc.
- Manual Nodos de modelado de IBM SPSS Modeler 14.2.

### Enlaces de interés:

- <ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/14.2/es/UsersGuide.pdf>
- [http://www2.microstrategy.com/producthelp/9.3/manuals/es/BasicReporting\\_Spanish.pdf](http://www2.microstrategy.com/producthelp/9.3/manuals/es/BasicReporting_Spanish.pdf)
- <http://www.R-project.org/>

### Herramientas utilizadas:

- MicroStrategy Desktop. Versión 9.2.1. Business Intelligence, Analytics & Mobile
- IBM SPSS Modeler. Versión 14.2.
- R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

## 8. LISTA DE FIGURAS Y TABLAS

Figura 1: Esquema de la recopilación de los datos de partida .....	6
Figura 2: Imagen de un ticket de compra .....	7
Figura 3: Imagen de un cuestionario de solicitud .....	7
Figura 4: Imagen de una Tarjeta de Fidelidad .....	8
Figura 5: Interfaz del sistema AS/400 .....	8
Figura 6: Estructura del Datawarehouse .....	10
Figura 7: Tabla tipos de datos .....	11
Figura 8: Modelo Entidad-Relación .....	11
Figura 9: Principales tablas de la base de datos .....	12
Figura 10: Estados o valores que pueden tomar los atributos .....	13
Figura 11 y 12: Comparativa atributo Sexo Asturias y Murcia VS Resto de España .....	16
Figura 13: Resultados Chi-Cuadrado para la variable Sexo .....	16
Figura 14: Tabla cálculo del contraste Chi-cuadrado para la variable Sexo .....	17
Figura 15 y 16: Comparativa atributo Tramo Edad Asturias y Murcia VS Resto de España .....	18
Figura 17: Resultados Chi-Cuadrado para la variable Tramo Edad .....	19
Figura 18: Tabla cálculo del contraste Chi-cuadrado para la variable Tramo Edad .....	20
Figura 19: Tabla desglose de la variable Segmento .....	21
Figura 20 y 21: Comparativa atributo Segmento y Murcia VS Resto de España .....	21
Figura 22: Resultados Chi-Cuadrado para la variable Segmento .....	22
Figura 23: Tabla cálculo del contraste Chi-cuadrado para la variable Segmento .....	22
Figura 24: Comparativa atributos VN y Op. en Asturias y Murcia VS Resto de España .....	24
Figura 25: Comparativa atributos MdC y TM en Asturias y Murcia VS Resto de España .....	25
Figura 26: Comparativa atributos Art. X Op. y Frec. en Asturias y Murcia VS Resto de España .....	25
Figura 27: Ruta SPDD Modeler para el cálculo de medidas centrales y de dispersión .....	26

Figura 28: Tabla medidas de tendencia central y de dispersión de las variables Art. X Op. y Frec. ....	26
Figura 29: Tabla resultados compras socios .....	28
Figura 30: Tabla ventas por familias de producto .....	30
Figura 31: Tabla de Producto para las cinco familias elegidas .....	31
Figura 32: Tabla de modelos de campaña .....	32
Figura 33: Modelo A Priori de asociación de familias .....	32
Figura 34: Gráfico de Mallas de familias de producto .....	33
Figura 35: Resultados Gráfico de Malla .....	33
Figura 36: Modelo de Árbol de Decisión C.5 .....	34
Figura 37: Pantalla principal del gestor de campañas Teradata Relationship Manager .....	36
Figura 38: Tabla cifras por canal de comunicación .....	36
Figura 39: Tabla de cálculo de los grupos de contro y de espejo .....	37
Figura 40: Tabla definición de las métricas calculadas para el análisis .....	38
Figura 41: Análisis general de la campaña en Asturias y Murcia .....	38
Figura 42: Resultado de análisis de producto .....	39
Figura 43: Resultados análisis por zona y canal de comunicación .....	39
Figura 44: Rentabilidad por canal de socios comunicados frente a su grupo de control .....	40
Figura 45: Rentabilidad de producto para la postal .....	41
Figura 46: Rentabilidad de producto para el sms .....	41
Figura 47: Rentabilidad de producto para el e-mail .....	42

# ANEXOS

# Anexo 1

Se va a explicar la fórmula utilizada para determinar el tamaño de la muestra que va a representar a la población, sobre la que se realizarán los estudios.

**POBLACIÓN.-** Llamado también universo o colectivo, es el conjunto de todos los elementos que tienen una característica común. Una población puede ser finita o infinita. Es población finita cuando está delimitada y conocemos el número que la integran, así por ejemplo, para este TFG: Socios con domicilio en las diferentes provincias de España. Es población infinita cuando a pesar de estar delimitada en el espacio, no se conoce el número de elementos que la integran.

**MUESTRA.-** La muestra es un subconjunto de la población. Ejemplo: Socios de Asturias y Murcia.

Sus principales características son:

**Representativa.-** Se refiere a que todos y cada uno de los elementos de la población tengan la misma oportunidad de ser tomados en cuenta para formar dicha muestra.

**Adecuada y válida.-** Se refiere a que la muestra debe ser obtenida de tal manera que permita establecer un mínimo de error posible respecto de la población.

Para que una muestra sea fiable, es necesario que su tamaño sea obtenido mediante procesos matemáticos que eliminen la incidencia del error.

## ELEMENTO O INDIVIDUO

Unidad mínima que compone una población. El elemento puede ser una entidad simple (una persona) o una entidad compleja (una familia), y se denomina unidad investigativa.

## FÓRMULA PARA CALCULAR EL TAMAÑO DE LA MUESTRA

Para calcular el tamaño de la muestra suele utilizarse la siguiente fórmula:

$$n = \frac{N\sigma^2 Z^2}{(N-1)e^2 + \sigma^2 Z^2}$$

Donde:

n = el tamaño de la muestra.

N = tamaño de la población.

$\sigma$  = Desviación estándar de la población que, generalmente cuando no se tiene su valor, suele utilizarse un valor constante de 0,5.

Z = Valor obtenido mediante niveles de confianza. Es un valor constante que, si no se tiene su valor, se lo toma en relación al 95% de confianza equivale a 1,96 (como más usual) o en relación al 99% de confianza equivale 2,58, valor que queda a criterio del investigador.

e = Límite aceptable de error muestral que, generalmente cuando no se tiene su valor, suele utilizarse un valor que varía entre el 1% (0,01) y 9% (0,09), valor que queda a criterio del encuestador.

La fórmula del tamaño de la muestra se obtiene de la fórmula para calcular la estimación del intervalo de confianza para la media, la cual es:

$$\bar{X} - Z \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \leq \mu \leq \bar{X} + Z \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

De donde el error es:

$$e = Z \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

De esta fórmula del error de la estimación del intervalo de confianza para la media se despeja la n, para lo cual se sigue el siguiente proceso:

Elevando al cuadrado a ambos miembros de la fórmula se obtiene:

$$(e)^2 = \left( Z \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \right)^2$$

$$e^2 = Z^2 \frac{\sigma^2 N - n}{n N - 1}$$

Multiplicando fracciones:

$$e^2 = \frac{Z^2 \sigma^2 (N - n)}{n(N - 1)}$$

Eliminando denominadores:

$$e^2 n(N - 1) = Z^2 \sigma^2 (N - n)$$

Eliminando paréntesis:

$$e^2 nN - e^2 n = Z^2 \sigma^2 N - Z^2 \sigma^2 n$$

Transponiendo n a la izquierda:

$$e^2 nN - e^2 n + Z^2 \sigma^2 n = Z^2 \sigma^2 N$$

Factor común de n:

$$n(e^2 N - e^2 + Z^2 \sigma^2) = Z^2 \sigma^2 N$$

Despejando n, y ordenando se obtiene la fórmula para calcular el tamaño de la muestra:

$$n = \frac{Z^2 \sigma^2 N}{e^2 N - e^2 + Z^2 \sigma^2} \quad \Rightarrow \quad n = \frac{Z^2 \sigma^2 N}{e^2 (N - 1) + Z^2 \sigma^2}$$

Esta fórmula se aplica tras suponer que los datos que se tienen se distribuyen normalmente; se llega a través de la definición que se tiene de *distribución normal*:

⇒ La distribución de probabilidad conocida como distribución normal, o campana de Gauss, es, por la cantidad de fenómenos que explica, la más importante de las distribuciones estadísticas. Multitud de variables aleatorias continuas siguen una distribución normal o aproximadamente normal. Una de sus características más importantes es que casi cualquier distribución de probabilidad, tanto discreta como continua, se puede aproximar por una normal.

La distribución de probabilidad normal y la curva normal que la representa, se caracteriza por:

- tener una forma de campana y un solo pico en el centro de la distribución. De esta manera, la media aritmética, la mediana y la moda de la distribución son iguales y se localizan en el pico. Así, la mitad del área bajo la curva se encuentra a la derecha de este punto central y la otra mitad está a la izquierda de dicho punto.
- los dos extremos de la distribución normal de probabilidad se extienden indefinidamente y nunca tocan el eje horizontal.

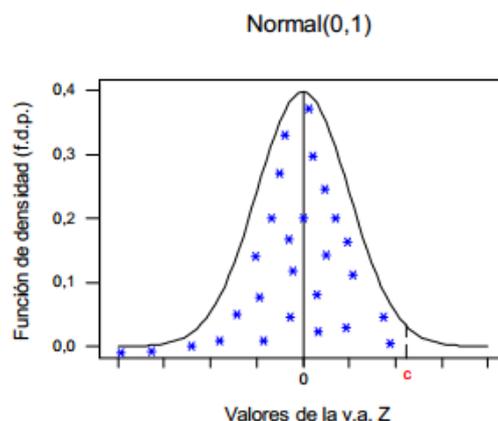
Para indicar que una variable aleatoria (v.a.) sigue una distribución normal de media  $\mu$  (o promedio de los datos) y desviación estándar  $\sigma$  (medida de cuánto se desvían los datos de su media), se usa la expresión:

$$X \sim N(\mu, \sigma)$$

La probabilidad de que una variable aleatoria (v.a.)  $X$  tome un valor determinado entre dos números reales  $a$  y  $b$  coincide con el área encerrada por la función de densidad de probabilidad entre los puntos  $a$  y  $b$ :

No existe una sola distribución de probabilidad normal; cada una de las distribuciones puede tener una media ( $\mu$ ) o una desviación estándar distinta ( $\sigma$ ). Por tanto, el número de distribuciones normales es ilimitado. La que se utiliza es la que tiene media es 0 y desviación estándar 1, conocida por distribución estándar normal. Todas las distribuciones normales pueden convertirse a la estándar, restando la media de cada observación y dividiendo por la desviación estándar.

Se convierte la distribución real en una distribución normal estándar utilizando un valor llamado  $Z$ , o estadístico  $Z$ :  $X \sim N(\mu, \sigma)$ , entonces la variable es  $Z = (X - \mu) / \sigma$  se distribuye según una normal de media 0 y desviación estándar 1,  $Z \sim N(0,1)$ , que es la distribución llamada normal estándar o tipificada.



## Anexo 2

El estadístico de la Chi-Cuadrado de Pearson o Ji-cuadrado ( $\chi^2$ ) es la prueba estadística no paramétrica por excelencia. Se basa en la comparación de frecuencias observadas y esperadas en una o más variables categóricas.

Se usa la prueba de Chi cuadrado cuando se quiere determinar si existe no una diferencia estadísticamente significativa entre la frecuencia observada y la frecuencia esperada que presentan los sujetos u objetos de una o más categorías. Pretende establecer si existe independencia o no entre variables categóricas.

Requisitos para llevar a cabo una prueba de Chi cuadrado:

1. Una base de datos cuantitativa.
2. Una o más categorías.
3. Observaciones independientes.
4. Un adecuado tamaño muestra (al menos 10).
5. Que se trate de una muestra aleatoria.
6. Que la base de datos se haya transformado a frecuencias.
7. Todas las observaciones deben ser usadas

La prueba de Chi-cuadrado parte del supuesto de que dos variables no están relacionadas, es decir, hay independencia. Por tanto se tienen dos hipótesis:

H0: Independencia de las variables.

H1: Variables relacionadas.

El estadístico:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

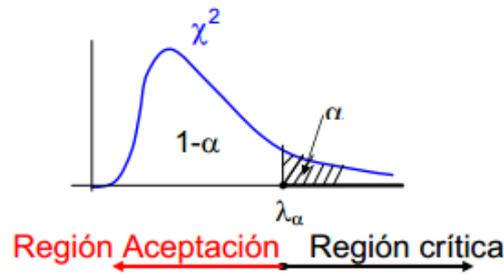
propuesto por Pearson, sigue una distribución chi-cuadrado con un número de grados de libertad igual al número de sucesos diferentes considerados, k, menos uno. Así, en el caso de  $GL = 2 - 1 = 1$ , que equivale a que en ese conjunto solo una variable ya conocida puede operar de manera independiente sin necesidad de que deba ser conocido el valor exacto de la otra.

**Nota.** El número de observaciones esperadas en cada clase debe ser mayor o igual a 5, es decir,  $E_j \geq 5$ . Si esto no ocurre se unen las clases adyacentes hasta cumplir el requisito. Al unir las clases se disminuirán los grados de libertad de la chi-cuadrado.

Este valor debemos compararlo con el valor crítico de la distribución  $\chi^2$  con  $(k-1)$  grados de libertad.

Cuanto mayor sea el valor de  $\chi^2$ , menos verosímil es que la hipótesis sea correcta. De la misma forma, cuanto más se aproxima a cero el valor de Chi-cuadrado, más ajustadas están ambas distribuciones.

La regla de decisión se observa en la figura.



El nivel de significación de la Chi-Cuadrado, representado por el símbolo  $\alpha$ , indica la probabilidad de cometer un error al rechazar la hipótesis nula, en el supuesto de que esta hipótesis es nula. Se representa por el símbolo  $\alpha$  y puede tomar dos valores, 0.01 y 0.05. Cuando el estadístico Chi-Cuadrado tiene una probabilidad menor a 0.05, se rechaza la hipótesis de independencia.

Tabla  $\chi^2$  de Pearson

Grados de libertad	$p = 0,05$	$p = 0,01$	$p = 0,001$
1	3,84	6,63	10,83
2	5,99	9,21	13,81
3	7,81	11,34	16,27
4	9,49	13,28	18,47
5	11,07	15,09	20,51
6	12,59	16,81	22,46
7	14,07	18,47	24,32
8	15,51	20,09	26,12
9	16,92	21,67	27,88
10	18,31	23,21	29,59
11	19,67	24,72	31,26
12	21,03	26,22	32,91
13	22,36	27,69	34,53
14	23,68	29,14	36,12
15	25,00	30,58	37,70
16	26,30	32,00	39,25
17	27,59	33,41	40,79
18	28,87	34,80	42,31
19	30,14	36,19	43,82
20	31,41	37,57	45,31
21	32,67	38,93	46,80
22	33,92	40,29	48,27
23	35,17	41,64	49,73
24	36,41	42,98	51,18
25	37,65	44,31	52,62
26	38,88	45,64	54,05
27	40,11	46,96	55,48
28	41,34	48,28	56,89
29	42,56	49,59	58,30
30	43,77	50,89	59,70
32	46,19	53,49	62,49

