



UNIVERSIDADE DA CORUÑA

Facultad de Economía y Empresa

Trabajo de
Fin de Grado

Análisis estadístico
aplicado: Obtención
de perfiles de
consumidores de
moda online

Nerea López Garrote

Tutor: Prof. Dr. Xose Manuel
Martínez Filgueira

Grado en Economía

Año 2017

Resumen

El objetivo de este trabajo es mostrar cómo se aplica la estadística a un ámbito más de la Economía, como es en este caso, a la investigación de mercados. El ejemplo que se ha realizado trata un tema, el comercio online, cada vez más presente en la sociedad.

En el trabajo se ha descrito el proceso y algunas herramientas estadísticas necesarias para una investigación de mercados, organizándolas en tres categorías: métodos univariantes, métodos bivariantes y métodos multivariantes (Trespalacios, Vázquez y Bello, 2005).

Siguiendo esta descripción, en la aplicación práctica se empieza recogiendo una muestra mediante una encuesta online. Una vez obtenida la muestra, el caso práctico se basa en analizar los perfiles sobre los consumidores de compra de ropa online, y a mayores también se observan otras variables como la satisfacción del cliente al realizar dicha compra o los motivos existentes de no realizar la compra de ropa por internet.

Puesto que el trabajo ya cuenta con una serie de limitaciones (de medios, de formato), no es posible explicar de forma detallada todo lo deseable e interesante de los datos obtenidos, con lo que se han escogido una serie de conclusiones relevantes, pero sin entrar a una mayor profundidad de análisis que hubiese sido posible al disponer de más medios.

Palabras clave: Estadística, Investigación de Mercados, Encuesta Web, Comercio electrónico.

Número de palabras: 13.756

Resumo

O obxectivo neste traballo é amosar como se aplica a estatística noutro ámbito máis da Economía, como é neste caso, á investigación de mercados. Como exemplo máis específico realizouse baixo un tema, o comercio electrónico, cada vez máis presente na sociedade.

No traballo describiuse o proceso e algunhas ferramentas estatísticas necesarias para unha investigación de mercados, organizándoas en tres categorías: métodos univariantes, métodos bivariantes e métodos multivariantes (Trespacios, Vázquez e Bello, 2005).

Seguindo esta descrición, na aplicación práctica comézase tomando unha mostra mediante enquisa online. Unha vez obtida a mostra, o caso práctico baséase en analizar os perfís dos consumidores de compra de roupa online e, a maiores, tamén se observan outras variables como a satisfacción do cliente ao realizar dita compra ou os motivos existentes de non realizar a compra de roupa por internet.

Posto que o traballo xa conta cunha serie de limitacións (de medios, de formato), non é posible explicar de forma detallada todo o que é desexable e interesante dos datos obtidos, co que se escolleron unha serie de conclusións relevantes, pero sen entrar a unha maior profundidade de análise que sería posible ao dispor de máis medios.

Palabras chave: Estatística, Investigación de Mercados, Enquisa Web, Comercio electrónico.

Número de palabras: 13.756

Abstract

The aim of this work is to show how statistics are applied to a wider field of economics, as in this case, to market research. The example that has been made deals with an issue, e-commerce, increasingly present in society.

The work has described the process and some statistical tools necessary for market research, organizing them into three categories: univariate methods, bivariate methods and multivariate methods (Trespalcios, Vázquez and Bello, 2005).

Following this description, the practical application begins by taking a sample through an online survey. Once the sample is obtained, the practical case is based on analyzing the profiles on the consumers of online clothing purchases, and also other variables are observed such as customer satisfaction when making such purchase or the existing reasons for not making the purchase of clothes by Internet.

Since the work already has a series of limitations (of means, of format), it is not possible to explain in detail everything that is desirable and interesting of the obtained data, with which have been chosen a series of relevant conclusions, But without going into a greater depth of analysis that would have been possible by having more means.

Keywords: Statistics, Market Research, Survey, E-Commerce.

Number of words: 13.756

Índice

Resumen	2
Resumo	3
Abstract.....	4
Índice de figuras	7
Índice de tablas.....	8
Introducción.....	9
1. La estadística.....	11
1.1 Definición y evolución de la Estadística.....	11
1.2 Conceptos estadísticos	13
1.2.1 Conceptos básicos	13
1.2.2 Elaboración de un análisis estadístico	15
1.2.3 Estadística y economía.....	16
2. Estadística e Investigación de Mercados	17
2.1 Introducción a la investigación de mercados.....	17
2.2 Métodos de recogida de datos en la investigación de mercados.....	19
2.2.1 Métodos cualitativos.....	19
2.2.2 Métodos cuantitativos	20
2.3 Métodos estadísticos aplicados a la investigación de mercados	26
2.3.1 Análisis Univariante	26
2.3.2 Análisis Bivariante	26
2.3.3 Análisis Multivariante	27
3. Aplicación práctica.....	30
3.1 Diseño previo.....	31
3.2 Análisis estadístico	32
3.2.1 Análisis Univariante: Estudio del perfil de los encuestados.....	33

3.2.2 Análisis Bivariante: Relaciones entre “compra” y perfil de los encuestados	36
3.2.3 Análisis Multivariante	47
3.3 Estudio sobre otros aspectos	54
Conclusiones	57
Bibliografía	59
Anexos	61
Anexo A Gráficos	61
Anexo B Tablas	62
Anexo C Cuestionario	64

Índice de figuras

Figura 1 Función de densidad de una Normal (0,1) (Campana de Gauss).....	15
Figura 2 Proceso de elaboración de un Análisis Estadístico	15
Figura 3 Actividades desarrolladas en una encuesta	22
Figura 4 Diagrama en frecuencias relativas de los niveles de compra de ropa online en función de la edad	37
Figura 5 Diagrama en frecuencias relativas de los niveles de compra de ropa online en función del género	38
Figura 6 Diagrama en frecuencias absolutas de los niveles de compra de ropa online en función de la ocupación	39
Figura 7 Diagrama en frecuencias absolutas de los niveles de compra de ropa online con respecto a los estudios	40
Figura 8 Diagrama en frecuencias relativas de los niveles de compra de ropa online en función de los ingresos individuales del sujeto	41
Figura 9 Diagrama de asociación de la compra de ropa online respecto a la edad	44
Figura 10 Diagrama de asociación de la compra de ropa online respecto a los estudios	45
Figura 11 Diagrama de asociación de la compra de ropa online respecto a la ocupación	46
Figura 12 Dendograma	50
Figura 13 Mapa perceptual de las dimensiones 1 y 2	53
Figura 14 Grado de satisfacción del cliente en frecuencias absolutas	55
Figura 15 Grado de satisfacción del cliente en frecuencias relativas	55
Figura 16 Diagrama de puntos de los usuarios que no compran en relación con sus motivos.....	56
Figura 17 Diagrama en frecuencias absolutas de los estudios en función de la edad	61
Figura 18 Diagrama en frecuencias absolutas de los ingresos en función de la edad	61

Índice de tablas

Tabla 1 Composición de la muestra por compra	33
Tabla 2 Composición de la muestra por género	34
Tabla 3 Composición de la muestra en función de la edad	34
Tabla 4 Composición de la muestra por ocupación	35
Tabla 5 Composición de la muestra por nivel de estudios.....	35
Tabla 6 Composición de la muestra por nivel de ingresos.....	36
Tabla 7 Contrastes Chi-Cuadrado y V de Cramer	43
Tabla 8 Distribuciones de la compra de ropa online con respecto a sus variables explicativas	51
Tabla 9 Valores de las variables sobre las cinco dimensiones ordenados bajo la Dim 2	54
Tabla 10 Coordenadas de las dimensiones.....	62
Tabla 11 Valores del contraste V de Cramer	63
Tabla 12 P-valores del contraste de Chi-Cuadrado (X^2)	63

Introducción

En el ámbito de la Economía, como en el conjunto de las ciencias sociales, la Estadística es una herramienta fundamental, cuyo manejo y conocimiento resulta imprescindible para un mejor desarrollo de los trabajos asociados a nuestra área de conocimiento.

Gracias al desarrollo tecnológico han surgido nuevas oportunidades, en forma de fuentes de datos o necesidades diferentes que la Estadística sabe afrontar. Con relación a esto, las nuevas tecnologías, en la actualidad, están presentes en cualquier aspecto de la vida. Desde realizar una simple compra online, como llevar a cabo la jornada laboral de un determinado agente a través del teletrabajo, como cualquier otro tipo actividad. Es por ello, que una de las cuestiones en las que se centra este trabajo es el mercado electrónico, ya que es realmente un auténtico fenómeno de amplia dimensión y el cual incide en cualquier sector empresarial.

En relación a esa cuestión, sin embargo, el objetivo principal escogido ha sido describir el conjunto de instrumentos estadísticos que se manejan en la Investigación de Mercados, mostrando su uso, precisamente, para el conocimiento de un mercado online, a partir de datos recogidos mediante una encuesta online.

El mercado escogido para el ejemplo práctico es el de la moda (en este caso se destaca la ropa). Y en el análisis se desea explicar, mediante una gran variedad de métodos estadísticos, las variables más significativas e importantes que influyen a la hora de que los clientes compren ropa a través del comercio electrónico.

El hilo conductor que sigue el trabajo en primer lugar son las explicaciones teóricas de la estadística e investigación de mercados; mediante los cuales se facilita la comprensión de la parte práctica del análisis.

La parte práctica, que será la segunda fase del trabajo, se compone de tres puntos, en un primer momento se hace una pequeña descripción de los datos y variables para el análisis. A continuación, se lleva a cabo el proceso estadístico, el cual se divide en tres fases: método univariante, método bivariante y método multivariante; aplicando

métodos que se encuadrarían dentro de esas categorías. Finalmente se realizará un proceso de síntesis de los resultados obtenidos.

El trabajo finaliza enumerando las diferentes conclusiones globales que se han obtenido en su realización.

1. La estadística

1.1 Definición y evolución de la Estadística

La Ciencia Estadística y su metodología constituyen, en la actualidad, una herramienta que ha trascendido mucho más allá de lo que originalmente dio lugar a su etimología (el estado), y son incontables las disciplinas en la que la utilizan. Y ello es así, a pesar de las múltiples críticas negativas que ha recibido a lo largo de su existencia, por ejemplo algunas de ellas reiteradamente citadas por Benjamín Disraeli, citado en Martín, Lafuente y Faura (2015), popularizada por Mark Twain “Hay mentiras pequeñas, grandes mentiras y estadísticas”. Con todo eso, su uso se ha ido incrementando debido a que la utilidad de la Estadística supera con creces sus posibles inconvenientes. De entre las disciplinas en las que se destaca la Ciencia Estadística, se encuentran las Ciencias Sociales, y más concretamente, aquellas que tienen una gran relación con el mundo de la empresa. Gracias a ella, se estudian fenómenos de los que generalmente, de lo que no se dispone una gran cantidad de datos (Martín, Lafuente y Faura, 2015).

La Estadística Descriptiva tiene su origen aproximadamente mil o dos miles años antes de Cristo, en Egipto, China y Mesopotamia, donde ya se hacían censos para la administración de los imperios. Pero es con la introducción de nuevas aplicaciones en la teoría de las probabilidades en el siglo XVIII la que funda la Estadística Matemática. El término de Estadística se debe posiblemente a G. Achenwall, conocido como el padre de la Estadística, profesor de la Universidad de Göttingen (Vargas, 1995).

Desde sus orígenes, la acepción vulgar del término Estadística hace referencia a toda una colección de datos numéricos presentada de forma ordenada. Autores como Martín Pliego (2004), afirman que la estadística no sólo se entiende como un conjunto de valores numéricos si no que el término estadística cuenta con otros posibles significados que se muestran a continuación.

El vocablo estadística, etimológicamente, procede del latino status y significa, situación en la que una persona o cosa y en general cada uno de los sucesivos modos de ser una persona o cosa sujeta a cambios que influyen en su condición. De ahí nace su primera vocación: constituirse como la exteriorización cuantitativa de las cosas de

Estado. Tan fácil, como imaginar que las sociedades humanas más primitivas estuviesen interesadas en enumerar sus características más relevantes (familias, utensilios de caza y labranza, cabezas de ganado, etc.) o como ya hemos mencionado anteriormente en la creación de los censos. (Martín Pliego, 2004).

Por otro lado Escuder y Santiago (1995) destaca dos acepciones del término (Estadística Descriptiva y Estadística Inferencial). Bajo la Estadística Descriptiva, se hace referencia a los censos o recuentos de poblaciones, de recursos naturales o industriales, o de cualquier otra manifestación del estado, comunidad, provincia, pueblo o clase. El análisis descriptivo, es un análisis empírico de datos. Se parte de un grupo de individuos o elementos cualesquiera, que se investiga y se intenta describir mediante una serie de técnicas propias de la Estadística Descriptiva. Se explica que las conclusiones obtenidas, son ciertas para dicho colectivo analizado, pero no pueden pormenorizarse a elementos individuales. La Estadística Descriptiva resume respuestas de grandes números de participantes en estadística simple y gracias a ello cuenta con una amplia facilidad de uso.

Una segunda definición es la Estadística Inferencial que tiene como estudio los hechos físicos o morales del mundo que se presentan a numeración o recuento y a la comparación de las cifras a ellos referente (Escuder y Santiago, 1995). El propósito primario de la Estadística Inferencial es emitir un juicio acerca de una población o la recolección total de todos los elementos acerca de los cuales el investigador busca información (William y Barry ,2009).

Es decir, se hace referencia a dos significados del término Estadística. Por un lado como una colección de datos numéricos presentados de forma ordenada, y otro lado como una ciencia o metodología, correspondientemente.

1.2 Conceptos estadísticos

1.2.1 Conceptos básicos

La Estadística se configura como la tecnología del método científico que proporciona instrumentos para la toma de decisiones cuando éstas se adoptan en un ambiente de incertidumbre, siempre que esta incertidumbre pueda ser medida en términos de probabilidad (Martín Pliego, 2004).

La Estadística es considerada como la aplicación del método científico en el análisis de datos, producidos a raíz de fenómenos no determinísticos, con el fin de obtener conclusiones racionales. Los fenómenos no determinísticos, son aquellos en los cuales el resultado se ve afectado por la aleatoriedad, es decir, aunque conozcamos todas las circunstancias en las que ocurre el fenómeno, no podemos indicar la certeza absoluta o el resultado con el que se producirá. En este caso, por ejemplo: el número de ventas realizadas cada mes por un vendedor se verá afectado por una subida (o bajada) de un valor bursátil un día determinado. Estos fenómenos también podrían tener una parte determinista, es decir, que no venga dado por una cierta aleatoriedad y pueda ser predecible. Pero dado que el resultado del fenómeno surge a raíz de combinar una parte determinista con una parte aleatoria, seguirá existiendo un cierto grado de incertidumbre en el resultado, por lo que se considerará globalmente como un fenómeno no determinista.

Hay que tener en cuenta que no se podrán realizar predicciones exactas sobre lo que ocurrirá con los fenómenos no deterministas. A pesar de ello, realizar este tipo de estadísticas, ayudará a indicar un grado de certeza con el que va a ocurrir el fenómeno o suceso; o determinar un rango de valores posibles indicando la confianza que tenemos de que ocurra el verdadero valor; o dar unos valores que resuman las pautas del comportamiento del fenómeno, y que observaríamos cuando el fenómeno se repita con frecuencia.

Puesto que el lenguaje que utiliza la Estadística es el matemático, el primer concepto que se debe asignar al realizar un análisis estadístico es la variable al carácter o fenómeno que se desea estudiar en la realidad. Sin embargo, no todos los caracteres que nos puedan llegar a interesar tienen concreciones de naturaleza cuantitativa, como puede ocurrir cuando investigamos por ejemplo: los partidos políticos, ausencia o

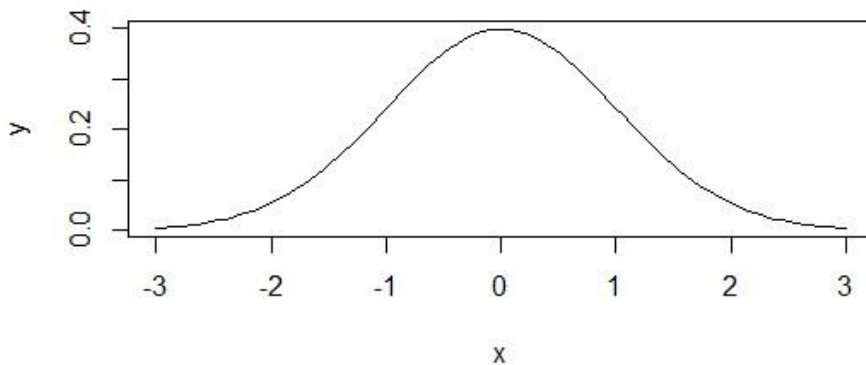
presencia, su sexo, las ramas de actividad económica, etc. A estas “variables cualitativas” se les designa con el nombre de atributos, ya que son variables que no aceptan medición (Ruiz-Maya, 2000).

Existen diferentes tipos de variables: las variables discretas y las variables continuas. Por un lado, la variable discreta es definida entre dos valores próximos, que puede tomar a lo sumo un número finito de valores (por ejemplo: número de hijos de una familia, número de clientes en una empresa, etc...). Y por otro lado, la variable continua, aquella que puede tomar infinitos valores de un intervalo (por ejemplo: el peso, la altura, entre otros). Esta distinción entre variables, es más teórica que práctica, puesto que la limitación de los aparatos de medida hace que casi todas las variables, prácticamente, se comporten como discretas cuando se pretenden observar.

Las variables de una población se ajustan en muchas ocasiones a modelos matemáticos preestablecidos. Siempre que la variable que se estudia se asocie a un modelo conocido. Al tener la fórmula matemática que siguen los datos se puede, por ejemplo, dibujar la curva que expresa ese modelo y aplicar los test estadísticos más específicos e idóneos para esas variables determinadas.

Existen modelos a los que se ajustan una gran cantidad de mediciones en la naturaleza y el modelo frecuente es el inicialmente descrito por A. Moivre y posteriormente desarrollado por C.F. Gauss que se conoce como modelo de distribución normal o de Gauss (Romero, 2015). Esta distribución cuenta con una función de densidad en forma de campana (véase **Figura 1**), simétrica respecto la media de la distribución. Para una distribución normal de media 0 y desviación típica 1. (William y Barry, 2009).

Figura 1 Función de densidad de una Normal (0,1) (Campana de Gauss)



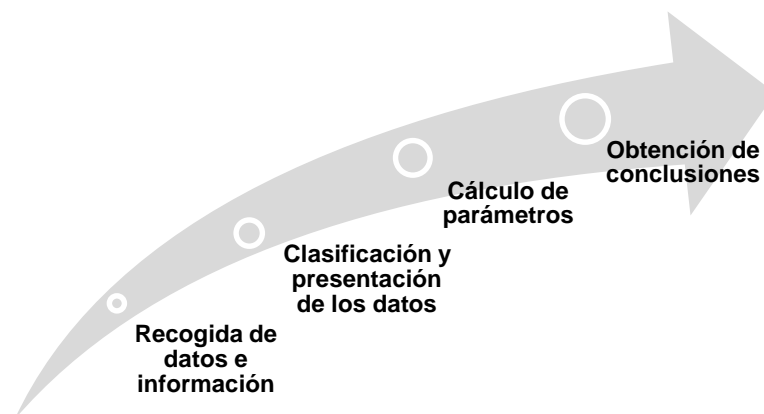
Fuente Elaboración propia en R

Otro aspecto importante a la hora de realizar una investigación estadística, es el conjunto de elementos que se somete a estudio, al que se le denomina población, puesto que ésta puede llegar a ser infinita o inabarcable, sólo será factible determinar una parte de esa población, a esta parte se le designa como muestra. El número de elementos que componen esa muestra, constituye el tamaño muestral y el proceso de selección de sus componentes es denominado como muestreo (Ruiz-Maya, 2000).

1.2.2 Elaboración de un análisis estadístico

Para llevar a cabo un análisis estadístico con un resultado satisfactorio se establecen una serie de etapas (véase **Figura 2**), habitualmente son las que se describen a continuación (Martín-Pliego, 2007):

Figura 2 Proceso de elaboración de un Análisis Estadístico



Fuente Elaboración propia a partir de Martín Pliego (2004)

- 1. Recogida de datos e información, que se lleva a cabo mediante la Teoría de muestras y el Diseño de experimentos. El planteamiento sobre lo que se puede investigar y la organización del trabajo de campo necesario para la recogida ha de diseñarse de manera minuciosa.
- 2. Clasificación y presentación de los datos, en la que se encontraría la Estadística Descriptiva. En general, se realiza organizando los datos en tablas simples o de doble entrada.
- 3. Cálculo de parámetros, que se llevaría a cabo mediante Inferencia Estadística.
- 4. Obtención de las conclusiones, a través del empleo de los métodos facilitados anteriormente, permitirá verificar las hipótesis sobre las regularidades que puedan detectarse.

1.2.3 Estadística y economía

Cabe destacar la importancia de la Estadística en la Economía ya que, la Ciencia Económica, como el resto de ciencias sociales, tiene por objeto el estudio del comportamiento del ser humano, y de la sociedad en general. Así, intenta determinar las interrelaciones y actitudes que poseen los individuos o grupos sociales ante los estímulos de carácter económico.

Por lo que la posibilidad de disponer de instrumentos objetivables, como los que la Estadística nos brinda, es fundamental para la verificación de las hipótesis que se plantean sobre un determinado comportamiento. Es más, constituye la única salida racional al proceso de investigación económica. El conjunto de técnicas estadísticas utilizables en la Ciencia Económica es, en parte, común a otras ciencias, y en parte, específico, ya que se han desarrollado métodos propios para el tratamiento de problemas específicos que han surgido en la estadística económica. Dentro de este campo más específico se citan por ejemplo: los análisis de concentración de la renta, el estudio de los números índices, buena parte del desarrollo estadístico de las series temporales, los modelos de contabilidad nacional y las tablas input-output de relaciones internacionales, los indicadores y los modelos regionales, las técnicas demográficas, etc. (Martín Pliego, 2004)

Además, se conocen una gran cantidad de situaciones en el mundo empresarial en las que surge la necesidad de una resolución mediante alguna técnica estadística. Por ejemplo, un analista financiero que desea determinar el estado económico de las empresas participantes en una industria. O un responsable del servicio de atención al cliente que desea identificar factores que diferencien de forma significativa entre los clientes que han realizado una devolución de sus compras y aquellos que no lo han hecho. O también, un analista de marketing que quiera conocer cuáles son los principales factores que influyen para que una empresa familiar tenga un comportamiento exportador, entre muchas otras situaciones comunes en los que la economía requiere de la estadística (Martín, Lafuente y Faura, 2015).

2. Estadística e Investigación de Mercados

2.1 Introducción a la investigación de mercados

En la actualidad, la búsqueda de información se ha convertido en un aspecto crucial para las empresas, generado en gran medida, por cambios que se producen constantemente en su entorno. Este entorno se ha vuelto cada vez más complejo, competitivo y sobre todo más incierto. Por ello la necesidad de información en cualquier empresa es cada vez mayor para la toma de decisiones futuras.

El interés de obtener información de calidad a la hora de decidir en el ámbito comercial de la empresa, es cada vez mayor. De ahí, que existan una gran cantidad de instrumentos que puedan favorecer la información de los agentes, es ahí donde se destaca, la investigación de mercados, ya que uno de sus principales objetivos, es obtener información que sea de ayuda en la toma de decisiones empresariales, logrando disminuir el riesgo de las mismas.

El Consejo de Directores de la American Marketing Association (AMA) en 2013 aclara que:

La investigación de mercados se define como la función que vincula al consumidor, cliente y público al vendedor a través de información que sirve para identificar y definir las oportunidades y los problemas de comercialización; generar y evaluar las actividades de marketing; supervisar el desempeño del marketing; y mejorar la comprensión del marketing como un proceso. En la investigación de mercados se especifica la información necesaria para abordar estos temas, se diseña el método de recogida de información, administra e implementa el proceso de recogida de datos, se analizan los resultados, y se comunica los resultados y sus implicaciones.

Algunos autores como Malhotra (2008), definen la investigación de mercados como la identificación, acopio, análisis, difusión y aprovechamiento sistemático y objetivo de la información con el fin de mejorar la toma de decisiones relacionada con la identificación y la solución de los problemas y las oportunidades de marketing. Conviene mencionar varios aspectos de esta definición.

En primer lugar, la investigación de mercados es sistemática; por lo que todas las etapas del proceso de investigación de mercados requieren una planeación metódica. Los procedimientos que se siguen en cada etapa son metodológicamente sólidos, están bien comprobados y, en la medida de lo posible, se planean de antemano.

Además, la investigación de mercados intenta aportar información precisa que refleje la situación real. Es objetiva y debe realizarse de forma imparcial. Aunque siempre es influida por la filosofía del investigador, que debería encontrarse libre de los sesgos personales o políticos del investigador.

La investigación de mercados incluye identificación, recopilación, análisis, difusión y uso de la información. Cada fase de este proceso es importante.

La puesta en marcha de una investigación de mercados es una tarea compleja que viene integrada en varias fases. Según Marqués (2015) se distinguen las siguientes fases:

- Fase I. Determinación de necesidades de información, es decir, es necesario concretar una serie de objetivos que serán la base para la investigación.
- Fase II. Diseño de la investigación de mercados, en la que se afronta una elección del tipo de investigación adecuada a las necesidades y se establece una serie de procedimientos para la selección de muestras de la población y los soportes que van a ser utilizados para captar la información requerida del mercado.
- Fase III. Ejecución de la investigación de mercados, en la que se lleva a cabo el trabajo de campo, es decir, se extraen los datos del mercado según la metodología elegida en la fase anterior. Además, se realiza un análisis con la información recolectada de forma que se obtengan resultados útiles para resolver el problema planteado.
- Fase IV. Comunicación de los resultados, es la última fase del proceso y en ella se llevan a cabo las funciones de dar a conocer e interpretar los resultados al cliente. Para ello, se elabora un informe que recoja de forma ordenada los detalles de la investigación realizada.

2.2 Métodos de recogida de datos en la investigación de mercados

La investigación de mercados cuenta con dos grandes vías de recogida de información, por un lado se encuentran, los métodos cualitativos y por otro los métodos cuantitativos. Se ha llegado a la conclusión que tanto los métodos cualitativos como los métodos cuantitativos se complementan, por lo que los dos juegan un papel esencial en el proceso de investigación de mercados.

2.2.1 Métodos cualitativos

Es este primer caso, los métodos cualitativos son aquellos que adquieren los conocimientos preliminares sobre los problemas y las oportunidades de decisión. Estos métodos presentan una serie de características comunes que se describen a continuación (Trespacios, Vázquez y Bello 2005):

- ✓ Proporcionan información de grupos reducidos de personas que no son estadísticamente representativas de la población a estudiar.
- ✓ Sus resultados se orientan a conocer y comprender opiniones, percepciones, actitudes, motivaciones y hábitos, es decir, centrándose más en por qué y cómo que en el cuantos.
- ✓ Metodológicamente son estudios muy flexibles, poco estructurados e informales.
- ✓ Su éxito depende de la creatividad, sentido común e intuición del investigador, aspectos de gran importancia para captar la naturaleza del problema en estudio. Por lo que, todos los puntos de vista son valiosos, evitándose los juicios de valor sobre las opiniones ajenas.

Siguiendo a los autores Trespalacios, Vázquez y Bello (2005) se encuentran las siguientes técnicas cualitativas: Dinámica de Grupo, Método Delphi, Técnicas de Creatividad, Técnicas proyectivas, Entrevista de pseudo-compra y Técnicas de observación.

2.2.2 Métodos cuantitativos

En los métodos cuantitativos de investigación de mercados, los datos recogidos sí que admiten algún tipo de medida y su análisis se realiza mediante procedimientos estadísticos, normalmente a partir de una muestra.

En este caso, la encuesta es el instrumento más utilizado en este tipo de investigación, puesto que ha mostrado ser un medio eficaz para conocer las características del comportamiento de los agentes y por ende, de una sociedad.

Pero además de la encuesta, otra opción como instrumento cuantitativo son los paneles de información, los cuales consisten en entrevistar periódicamente a una muestra representativa de la población con la finalidad de obtener, además de las características de los encuestados, información sobre sus comportamientos en un horizonte temporal a largo plazo. De este modo, se podría conseguir un conocimiento

de las decisiones futuras del consumidor, diferenciándose así, de la encuesta, que cuenta con una visión más estática, más fija (Trespalacios, Vázquez y Bello, 2005).

Este trabajo se centra en la Estadística aplicada en la Investigación de Mercados y en el que se ha escogido como método de recogida de información, la encuesta. Por lo tanto, el inicio del análisis se centra en el análisis de esa encuesta, que es la primera de las técnicas estadísticas que se abordan.

Una de las principales ventajas de la encuesta viene otorgada por que nos permite recoger una gran cantidad de información con garantías estadísticas, que se pueden extrapolar a toda una población¹. Este método de recogida de la información hace posible la realización de un análisis estadístico que logrará explicar y comprender los posibles conceptos y relaciones que existan entre las variables.

La encuesta presenta una serie de ventajas por las que llevarla a cabo (Abascal y Grande, 2005) como pueden ser:

- La estandarización. La encuesta permite hacer siempre las mismas preguntas a todos los elementos de la muestra, esto implica una homogeneidad de la información.
- Facilidad de administración. Con este método se obtiene información a partir de un cuestionario en el que se leen los elementos de la muestra.
- Simplificación del tratamiento de datos. Las encuestas poseen números y códigos que identifican las respuestas, es decir, se pueden codificar la encuesta, o convertir a números aspectos cualitativos. Esto facilita grabar los datos en ordenadores para poder tratarlos posteriormente en una infinidad de programas.
- Obtención de información no directamente observable. Al llevar a cabo una encuesta se pueden conseguir información relativa a los encuestados como el aspecto, su hábitat, etc.
- Posibilidad de hacer estudios parciales. Debido a que las encuestas cuentan con identificadores de características de las personas que

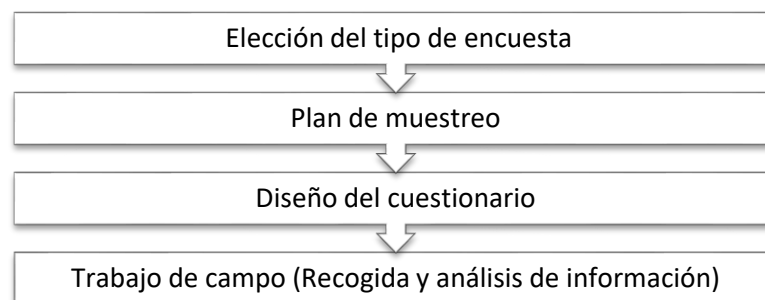
¹ Población: es la suma de todos los elementos que comparten un conjunto común de características y que constituyen el universo para el propósito del problema de la investigación de mercados (Malhotra, 2008).

responden, es posible hacer estudios imponiendo condiciones. Por ejemplo, se puede estudiar la intención de voto de una población atendiendo a su edad, género, zona, etc.

A pesar de sus ventajas, la encuesta también se encuentra con una serie de limitaciones. Como la existencia de determinados aspectos y actitudes de los seres humanos que no se pueden medir o captar mediante una serie de preguntas en forma de cuestionario. Otro inconveniente, es que pueden estar mal formuladas, o con un planteamiento deficiente que podría otorgar una falsa imagen de lo que sucede en la realidad. Y en el mismo sentido actuarían el diseño de preguntas confusas o ambiguas que dificultarían el trabajo realizado por este método (Trespalcios, Vázquez y Bello, 2005).

Por ello es esencial determinar una serie de etapas para realizar de forma correcta una encuesta, como se observa en la **Figura 3**, en un primer momento la primera fase es la elección del tipo de encuesta, seguido de una definición del plan de muestreo, el diseño el cuestionario y la realización del trabajo de campo con la información recogida.

Figura 3 Actividades desarrolladas en una encuesta



Fuente Elaboración propia a partir de Trespalcios, Vázquez y Bello (2005)

La elección del tipo de encuesta es la premisa inicial a la hora de elaborar una investigación de mercados utilizando la encuesta como método de recogida de información. Existen varias opciones de encuestas: encuestas telefónicas, encuestas personales, encuestas por correo y encuestas electrónicas. De todas estas técnicas, las

encuestas telefónicas son las más populares, seguidas de las encuestas personales y de las encuestas por correo. Pero el uso de recursos electrónicos, facilita enormemente las encuestas por Internet por lo que éstas crecen a un ritmo acelerado en la actualidad (Malhotra, 2008).

Factores como la necesidad de información, las limitaciones del presupuesto (tiempo y dinero) y las características de los encuestados marcarán los resultados a la hora de elegir entre ninguno, uno, dos o incluso todos los tipos de cuestionarios adecuados a la investigación. Las diversas formas de recolección de datos no son mutuamente excluyentes, sino que se pueden utilizar de manera complementaria para sumar sus ventajas y compensar sus debilidades

A continuación, se determinará el plan de muestreo que se ha de llevar a cabo. Elegir una técnica de muestreo implica muchas decisiones de amplia naturaleza. El investigador debe decidir si usará una técnica de muestreo tradicional o una técnica bayesiana, si obtendrá la muestra con o sin reemplazo y si utilizará un muestreo probabilístico o no probabilístico.

En un primer momento se distingue entre técnica de muestreo tradicional versus técnica de muestreo bayesiano según aclara Malhotra (2008). La técnica de muestreo tradicional selecciona toda la muestra antes de iniciar la recolección de datos. Mientras que por lo contrario, la técnica bayesiana selecciona de manera secuenciada cada elemento, después cada elemento lo agrega a la muestra, se reúnen los datos, se calculan los estadísticos de la muestra y se determinan los costes de la muestra. Es por ello que ésta última no es común en la investigación de mercados ya que no se dispone de mucha de la información requerida sobre costes y probabilidades.

Pero sin duda la decisión más importante en la fase de plan de muestreo es la elección de la técnica de muestreo adecuada a cada investigación, que concierne al uso del muestreo probabilístico o no probabilístico.

Los procedimientos no probabilísticos, en los que las unidades muestrales no se seleccionan al azar sino que son elegidas por el investigador. Entre ellos podemos distinguir varios tipos:

- Muestreo por conveniencia. Los individuos son elegidos para formar parte de la muestra con un objetivo en concreto.
- Muestreo por juicios. (causal o incidental). El investigador selecciona directa o intencionadamente los individuos de la población, por ejemplo, selecciona a los que tiene más fácil acceso.
- Muestreo por cuotas. Se basa en que se conocen los individuos más adecuados para la investigación y se fijan unas cuotas que consisten en un número determinado de elementos que reúnen unas determinadas condiciones. Suele ser empleado en las encuestas de opinión.
- Muestreo por bola de nieve. Se eligen individuos, los cuales conducen a otros y estos a otros, hasta conseguir la muestra.

En segundo lugar, los muestreos probabilísticos. En ellos todos los elementos de la población tienen la misma probabilidad de ser elegidos. De esta forma se asegura que las características importantes de la población están representadas en la proporción que le corresponde. La selección se lleva a cabo de forma mecánica, y no tiene ningún carácter subjetivo con referencia al investigador. Distinguimos:

- Muestreo aleatorio simple. Se eligen a los individuos de forma aleatoria. Es aconsejable para las poblaciones de tamaño pequeño.
- Muestreo aleatorio sistemático. Es útil para poblaciones pequeñas y se realiza escogiendo un número al azar.
- Muestreo aleatorio estratificado. En este caso, más aconsejado para poblaciones grandes. Se divide la población en estratos según alguna de las características que queramos analizar (ejemplo: grupos de edad, situación profesional, etc.) y se intenta que en la muestra esté representado cada estrato en la proporción que le corresponde. Dentro de cada estrato se eligen a los individuos de forma aleatoria.
- Muestreo por conglomerados. O más conocido como *Análisis Cluster*, cuenta con la peculiaridad ya que su finalidad es dividir en grupos un conjunto de elementos que están definidos por diferentes variables, de forma que los elementos de un mismo grupo sean muy similares entre sí y los de grupos diferentes sean distintos.

No se debe olvidar que otra de las partes principales y más importantes en la elaboración de una encuesta es el diseño del cuestionario, ya que en él se recogen un conjunto de preguntas sobre las variables a investigar, que serán la clave para la obtención de unos resultados exitosos.

Todo cuestionario tiene tres objetivos específicos (Malhotra, 2008). Primero, debe traducir la información necesaria en un conjunto de preguntas específicas que los encuestados puedan responder. Es difícil desarrollar preguntas que los encuestados puedan y quieran responder y que brinden la información deseada. Dos formas aparentemente similares de plantear una pregunta pueden obtener información diferente. Por lo tanto, este objetivo es todo un reto.

Segundo, el cuestionario debe animar, motivar y alentar al encuestado para que participe activamente en la entrevista, colabore y concluya el proceso. Al diseñar un cuestionario, el investigador debe esforzarse por minimizar la fatiga, el aburrimiento, la falta de interés o la ausencia de respuestas por parte del encuestado.

Tercero, el cuestionario debe minimizar el error de respuesta. El error de respuesta es el que surge cuando los encuestados dan respuestas incorrectas, o cuando sus respuestas se registran o se analizan mal.

Además, para que estas preguntas se transformen en datos interesantes se han de poder medir y valorar, es decir, los datos han de ser operativos.

Con respecto a los datos es también importante tener en cuenta sus propiedades, en función de las cuales se clasifican en una escala de cuatro categorías (nominal, ordinal, de intervalo y de razón) propuestas por S. S. Stevens (1946), las cuales, se pueden resumir en dos: por un lado las escalas no métricas, propias de las variables cualitativas, que resultan más sencillas y las escalas métricas propias de las variables cuantitativas y con mayor posibilidad de análisis estadístico.

Una vez realizado el cuestionario, se ha de proceder a un análisis con el fin de leer, explotar e interpretar los datos disponibles y extraer de ellos información de interés para la investigación (trabajo de campo). Este análisis cuenta con diferentes clasificaciones, en este caso, se destaca la clasificación según el número de variables

empleadas para el análisis (univariable, bivariable y multivariable) según Trespalacios, Vázquez y Bello (2005).

2.3 Métodos estadísticos aplicados a la investigación de mercados

2.3.1 Análisis Univariante

El primer caso, **el análisis univariante**, es el caso más simple y sencillo, ya que cada variable de la matriz de datos se considera por separado. Intentando resumir o comprender mejor las características de la variable. Permiten una primera visión de las características estudiada.

- En el aspecto descriptivo del análisis, las fórmulas que se usan para las variables no métricas son: *la moda y las tablas de frecuencia* .Y para las variables métricas se hace uso de *la media aritmética y la desviación típica*.
- Mientras que en el análisis inferencial se emplean para las variables métricas se hace uso de *contrastos para diferentes parámetros*, como la media aritmética o la varianza.

2.3.2 Análisis Bivariante

El análisis bivariante, se observa cuando de forma conjunta en el análisis estadístico se consideran variables de dos en dos. Se trata de comprobar la posible existencia de relación entre ambas variables, es decir, si muestran o no dependencia.

En marketing es habitual llevar a cabo análisis de este tipo, estudiando si existe alguna relación entre cualquier variable de clasificación del consumidor y su comportamiento o actitud.

Por ejemplo, un investigador de una encuesta puede estar interesado en si los consumidores masculinos o femeninos adquieren un producto en la misma cantidad. (William y Barry ,2009).

Las fórmulas estadísticas que se utilizan en este caso son las siguientes:

- *La tabla cruzada o tabla de contingencia*, es la técnica empleada con variables no métricas para comprobar la existencia de relación entre dos variables.

Se ubica entre las técnicas estadísticas de mayor uso entre los investigadores de mercados. Las tabulaciones cruzadas son intuitivas y de fácil comprensión. También se prestan a un análisis gráfico con herramientas como los diagramas de barras.

El contraste de independencia de X^2 proporciona un medio para probar el significado estadístico de la tabla de contingencia. En otras palabras, la prueba X^2 bivariada examina el significado estadístico de relaciones entre dos variables.

- *La correlación lineal y el test de medias* son las técnicas más usadas con variables métricas. Con ellas se comprueba la posible existencia de relación entre dos variables y se contrasta si la diferencia entre dos medias muestrales infieren diferencias en las medias de los segmentos de la población, correspondientemente.

2.3.3 Análisis Multivariante

Sin embargo, cuando en el análisis tienen en cuenta, tres o más variables, las **técnicas se denominan multivariante**, este análisis se determina como el corazón de la investigación de mercados, ya que sirve para modelizar prácticamente todos sus eventos (Marqués, 2015). Se dividen en dos grupos principales: las técnicas de dependencia (o técnicas predictivas) y las técnicas de interdependencia (o técnicas descriptivas).

En el primer caso, las técnicas de dependencia, las cuales especifican el modelo para los datos en base a un conocimiento teórico previo. El modelo supuesto para los datos debe contrastarse después del proceso de ajuste de datos a un modelo estimado antes de aceptarlo como válido. Se incluyen las técnicas de: todo tipo de regresiones, análisis de varianza y covarianza, análisis discriminante y series temporales.

Por otro lado, en las técnicas de interdependencia no se asigna ningún papel determinado a las variables. No supone la existencia de variables dependientes ni independientes y tampoco necesita un modelo previo para los datos.

En este caso se encuentran *técnicas de reducción de la dimensión* (factorial, componentes principales, análisis de correspondencias, etc.), *técnicas de escalamiento óptimo y multidimensional* y *el análisis conjunto*. Además también se observan las técnicas de clasificación que extraen perfiles de comportamiento o clases, a lo que se le conoce como: *análisis clúster*.

De entre estos métodos se van a explicar con algo de detalle dos de ellos: el análisis cluster y el análisis de correspondencias, ya que se van a emplear para analizar los datos en la aplicación práctica.

Análisis Cluster

El análisis cluster es un conjunto de métodos o técnicas estadísticas que permiten describir y reconocer diferentes agrupaciones que subyacen en un conjunto de datos o unidades de observación, es decir, permiten clasificar o dividir en grupos más o menos homogéneos, un conjunto de individuos que están definidos por diferentes variables. Y el objetivo principal del análisis de conglomerados consiste, por tanto, en conseguir una o más particiones de un conjunto de individuos en base a determinadas características de los mismos (Pérez, 2007).

La homogeneidad o cercanía se mide entre los individuos para formar parte de un grupo de individuos que se miden en función de las distancias, preferentemente, aunque pueden usarse en diferentes situaciones otras medidas, como coeficientes de correlación o de asociación. La estrategia más común consiste en medir la equivalencia en términos de la distancia entre los pares de objetos. Los objetos con distancias reducidas entre ellos son más parecidos entre si que aquellos que tienen distancias mayores y se agruparan, por tanto, dentro del mismo cluster.

Además de las diferentes medidas de cercanía, también existen diferentes formas de llevar a cabo la agrupación de individuos, es decir, dependiendo del algoritmo que se utilice para llevar a cabo la agrupación de individuos, se obtienen diferentes métodos de análisis de conglomerados. La clasificación más habitual para estos métodos

establece una distinción entre los análisis cluster jerárquico y cluster no jerárquico (Pérez, 2007).

Un análisis cluster jerárquico tiene una estructura arborescente de dependencia que parte de tantos grupos iniciales como individuos se estudien y el objetivo se centra en conseguir agrupaciones sucesivas entre ellos de forma que progresivamente se juntarán hasta llegar al cluster final que contiene todos los casos analizados. Esta estructura arborescente está asociada a una representación gráfica denominada dendograma. Una cuestión importante en los métodos jerárquicos es decidir el número de grupos diferentes con los que se queda el investigador. Habitualmente esta decisión se toma cuando las diferencias entre los clusters agrupados son comparativamente grandes en relación con las agrupaciones anteriores.

Por lo contrario, los métodos de análisis cluster no jerárquicos, se diferencian de los métodos jerárquicos en que el investigador debe especificar a priori los grupos que deben ser formados y que trabajan con la matriz de datos original. Los individuos se asignaran a cada uno de esos grupos en función de diferentes algoritmos aplicados.

Análisis de Correspondencias

El análisis de correspondencias es un método multivariante de reducción de la dimensión. Autores como Clausen (1998) afirman que, el Análisis de Correspondencias (AC) es una herramienta que permite analizar la asociación entre dos o más variables categóricas representando las categorías de las variables como puntos en un espacio de pocas dimensiones.

Como describe Fuente (2011), el Análisis de Correspondencias tiene dos objetivos básicos: encontrar asociación entre categorías de una misma variable o encontrar la asociación entre categorías de variables distintas. Además, permite analizar cómo está estructurada esta asociación, indicando en qué grado contribuyen a esa relación los distintos valores de las variables, proyectando sobre un mapa perceptual toda esa información.

Existen dos métodos dentro del AC. El Análisis de Correspondencias Simple (ACS) que estudia el comportamiento conjunto de, únicamente, dos variables

cualitativas y el Análisis de Correspondencias Múltiple (ACM) cuando se analizan más de dos variables simultáneamente.

El ACS según Fuente (2011) analiza conjuntamente el comportamiento de dos variables partiendo de una tabla de contingencia. Las categorías de cada una de las variables figuran en filas y en columnas, respectivamente en la tabla. Su objetivo es simplificar esas relaciones, usando la asociación existente entre las diferentes categorías para representarlas en un eje de coordenadas, preferiblemente con un máximo de dos dimensiones. Para conseguir este objetivo crea nuevas variables independientes entre sí que serán los ejes o dimensiones en la representación gráfica, y en los que se intenta conservar la mayor cantidad posible de la información inicialmente contenida en los datos.

Cuando el número de variables es mayor que dos se trata de un ACM. Es el análisis más frecuente en investigación. En este caso se utiliza una tabla más compleja denominada Tabla de Burt (Fuente, 2011). Sin embargo el objetivo es el mismo, reducir la complejidad de las categorías de todas las variables en un número menor de variables independientes denominadas dimensiones.

La representación gráfica tanto del ACS como del ACM permite una visualización de las relaciones entre las categorías de las variables estudiadas. La proximidad o distancia entre ellas y diferentes atributos como la contribución de cada variable en la formación de dimensiones permite explicar unos determinados perfiles de relación.

3. Aplicación práctica

Como muestra del trabajo estadístico asociado a la investigación de mercados se ha realizado una aplicación práctica, en la que se ha recogido mediante una encuesta online, la información sobre la opinión de un grupo de consumidores sobre el uso de webs de moda para la compra de ropa online.

En ella se muestran los diferentes aspectos descritos en la primera parte de este trabajo, comenzando con el diseño del cuestionario, recogida de datos, realización de diferentes análisis estadísticos y obtención de conclusiones.

3.1 Diseño previo

La Organización Mundial del Comercio en 1998, define el comercio electrónico como “la producción, distribución, comercialización, venta o entrega de bienes y servicios por medios electrónicos”.

En la actualidad, Internet es un auténtico fenómeno de dimensión mundial con incidencia en todos los sectores empresariales, entre ellos el de la moda. Su función es ayudar a comunicar, y es a través de esta comunicación como las empresas del sector de la moda pueden informar en tiempo real y a un bajo coste sobre cualquier oferta de productos, presentar colecciones, etc. Cada vez son más las empresas que invierten en este medio con el objetivo de mejorar los canales de venta e incrementar sus ingresos (Olmo Arriaga y Fondevila, 2009).

Esta es una de las razones por las que en la parte práctica se elaborará un ejemplo relacionado con el consumo de moda a través de internet. Además, otro motivo es que la moda es algo tan versátil como la vida misma, es un factor psicológico sumergido en la mente de las personas, y que personalmente me fascina. Es por ello que en este análisis el objetivo será determinar cuáles son los factores o variables relevantes a la hora de identificar un cliente o posible cliente en compra de ropa online. Una matización sobre el alcance de este análisis es que, debido a las limitaciones de medios y de extensión, se ha marcado como objetivo el manejo de las diferentes técnicas estadísticas comentadas, pero sin entrar a fondo en su aplicación.

Como método de recogida se ha llevado a cabo una encuesta online, debido a las ventajas que nos proporciona para información con garantías rapidez y a bajo costo.

La encuesta por Internet tiene grandes virtudes como la gran rapidez, la mejora en la respuesta por la posibilidad de introducir elementos audiovisuales en el cuestionario y el menor coste de la investigación cuando se compara con encuestas. Pero al mismo

tiempo no se obvian los inconvenientes como los errores de cobertura o la ausencia de aleatoriedad de las muestras y la menor tasa de respuesta (Vidal, 2012).

A continuación, se define cuáles son los objetivos de la encuesta, que estarán determinados por los objetivos de la investigación que se desea estudiar, como pueden ser: comportamientos, actitudes u opiniones (Abascal y Grande, 2005). En este caso lo que se recoge es la opinión de un grupo de consumidores sobre la compra de ropa online. A partir de ese diseño se realiza el de cuestionario que se lleva a cabo en internet a partir de un cuestionario de la plataforma Google. En este cuestionario se establecen la serie de preguntas que darán lugar a nuestra muestra. El cuestionario se observa en el **Anexo C** o mediante el siguiente enlace:

<https://goo.gl/forms/yH8Xt7NzNM86uPzR2>

Los datos extraídos se gestionan con un programa estadístico llamado R, el cual es un paquete que se sitúa entre un lenguaje y entorno de programación, y un software para realizar análisis estadísticos y generar gráficos. Fue desarrollado por Ihaka y Gentleman (1997) del Departamento de Estadística de la Universidad de Auckland (Nueva Zelanda).

3.2 Análisis estadístico

El estudio se ha llevado a cabo sobre una muestra en la que los sujetos, exactamente 430, han participado en la encuesta, por lo que no es una muestra aleatoria.

Para el estudio se han seleccionado cinco variables (explicativas) que describen el perfil de las personas participantes, y cuatro variables (explicadas) que tratan la existencia o no de compra de ropa online.

Antes de dar comienzo al análisis de los datos se ha seguido un proceso de depuración de esos mismos, puesto que se han descartado inmediatamente de los datos las respuestas en las que se pudiesen observar posibles fallos de los encuestados a la hora de contestar, como pueden ser: respuestas de los individuos en instantes de tiempo muy seguidos². También se ha eliminado una respuesta incoherente como ha

² Se producen cuando una persona no tiene claro que se haya guardado el formulario y vuelve a intentarlo

sido: una mujer de 55 años que establecía estudiar, no compraba ropa por internet pero su nivel de satisfacción era alto. Con lo cual once han sido el número total de respuestas eliminadas. Por ello, la muestra se ha reducido a un total de 419 individuos.

La mayor parte del análisis se realizará sobre los perfiles de compra, o sea, el análisis de las diferentes frecuencias de compra y las características de las personas que las realizan. Más brevemente, se analizan también los niveles de satisfacción con la compra de las personas compradoras y los motivos que aducen las personas que no han comprado.

3.2.1 Análisis Univariante: Estudio del perfil de los encuestados

En un primer instante, se lleva a cabo un estudio descriptivo de los datos (primera aproximación a los datos utilizados en el estudio y proceso de familiarización con los mismos), ello responde a la descripción de las variables determinantes y al perfil del investigado en cuestión. Se observa la variable explicada, “*compra*”, y las cinco variables explicativas que son: “*edad*”, “*género*”, “*ocupación*”, “*estudios*” e “*ingresos*”.

Compra: En la muestra recogida, como se puede observar en la **Tabla 1**, más de la mitad de los sujetos sí que compran por Internet aunque se dividen entre los niveles por intensidad de compra que son los siguientes: “*sí, con frecuencia*”; “*sí, ocasionalmente*” y “*muy poco (1 o 2 ocasiones)*”. Exactamente 295 (68,60%) sujetos sí que comprarían, frente a un (31,39%) 135 que no lo harían.

Tabla 1 Composición de la muestra por compra

Si, con frecuencia	57
Si, ocasionalmente	123
Muy poco (1 o 2 ocasiones)	103
No	135

Fuente Elaboración propia en R

Esto genera desde un primer momento una motivación más en el análisis, puesto que surgen una variedad de preguntas como: ¿Quiénes compran más? .Y lo más importante, ¿Por qué compran más?, ¿Hay alguna característica en común entre los sujetos que motive esa compra online?

A continuación se comentan las variables explicativas con los datos correspondientes que se utilizan en este análisis:

Género: La primera variable explicativa se centra en el sexo de estos sujetos como se puede observar en la **Tabla 2**, se destaca que un 70.64% de mujeres han contribuido a la muestra y por lo contrario, los hombres han contribuido en un 29.36%.

Se confirma aquí un sesgo en la muestra ya que no mantiene las proporciones reales de la población. Esto, que aquí es obvio, puede aparecer sin ser detectado en todas las variables estudiadas, por lo cual todas las conclusiones obtenidas serán simplemente sobre la muestra y, aunque se considere que pueden proporcionar información sobre pautas de consumo, no se pueden aceptar con la contundencia que tendrían si la recogida de la muestra fuese con un método probabilístico.

Tabla 2 Composición de la muestra por género

Hombre	123
Mujer	296

Fuente Elaboración propia en R

Edad: La siguiente variable es la edad de los encuestados que se observa en la **Tabla 3**, el rasgo más importante de esta variable es la gran cantidad de individuos menores de 25 años, exactamente un 40.28%. Este rango va a influir en las demás variables explicativas, puesto que “*hasta 25 años*” va a implicar un alto número de estudiantes, y por lo tanto de personas con ingresos bajos. Otra característica a destacar es la poca respuesta en el grupo de edad “*más de 55 años*” sólo tiene un 6.23% de las respuestas.

Tabla 3 Composición de la muestra en función de la edad

Hasta 25 años	168
Entre 26 y 35 años	58
Entre 36 y 45 años	99
Entre 46 y 55 años	66
Más de 55 años	26

Fuente Elaboración propia en R

Ocupación: Por otro lado, respecto la ocupación de los agentes como se muestra en la **Tabla 4**, se puede ver que más de la mitad de encuestados está trabajando, exactamente 222 sujetos que representan un 52.82% de la muestra. Seguido de 142 personas (34.05%) los cuales se encuentran estudiando. Los demás niveles de esta clasificación son más minoritarios y seguirían respectivamente el orden siguiente según sus datos: “en paro”, “labores de casa” y “jubilados”.

Tabla 4 Composición de la muestra por ocupación

Trabajando	222
Labores de casa	18
Estudiando	142
En paro	20
Jubilado	15

Fuente Elaboración propia en R

Estudios: En base a los estudios (**Tabla 5**), la mayoría de los encuestados cuentan con estudios universitarios (46.67%). El otro grupo notable, son los agentes que cuentan con formación profesional (21.42%). Del mismo modo y de forma minoritaria se encuentran los sujetos que han contestado a la muestra que tienen primaria o secundaria, con un 16.19% y 14.76%, respectivamente.

Tabla 5 Composición de la muestra por nivel de estudios

Primaria - EGB - ESO	68
Secundaria - BUP - Bachillerato	62
Formación Profesional	90
Estudios Universitarios	196

Fuente Elaboración propia en R

Ingresos: Finalmente se encuentra la variable vinculada a la capacidad económica del individuo (**Tabla 6**). En este caso destaca un 34.76% de los encuestados que ingresan menos de 600€ mensuales, resultado relacionado con el alto número de sujetos menores de 25 años y el alto número de estudiantes de la muestra. Un 19.52% de los sujetos entrevistados cuentan con un ingreso mensual de entre 1000€ y 1499€. Seguidos de agentes que obtienen entre 600€ y 999€ que representan un 18.80%. Mientras que por último se encontrarían los encuestados que ganan más de 2000 € y los de entre 1500 € y 1999 € con un 7.14% y 6.19%, correspondientemente.

Tabla 6 Composición de la muestra por nivel de ingresos

Menos de 600	146
Entre 600 - 999	79
Entre 1000 - 1499	82
Entre 1500 - 1999	26
Más de 2000	30

Fuente Elaboración propia en R

Después de este primer análisis se tiene un perfil del conjunto de la muestra recogida, y por lo tanto de sus diferencias con la población real producidas por su falta de aleatoriedad. Se puede destacar de este perfil el mayor número de mujeres en la muestra, quizás producido por el tema escogido. Otro aspecto relevante está en la edad, alto número de menores de 25 años y bajo de mayores de 55 años, lo que condicionará también la proporción en otras variables.

3.2.2 Análisis Bivariante: Relaciones entre “compra” y perfil de los encuestados

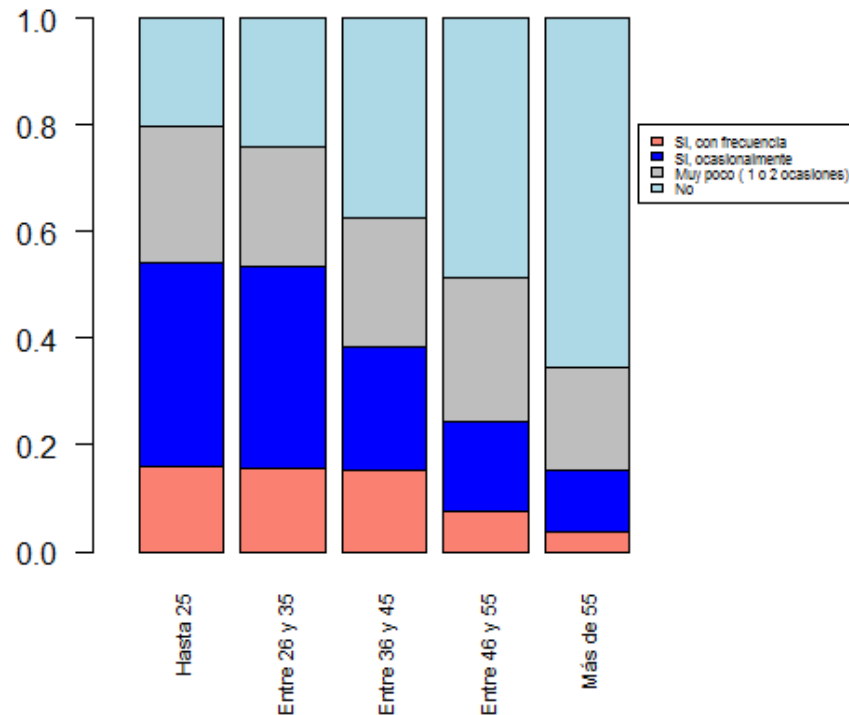
En esta parte del análisis se va a comparar las respuestas de la variable “*compra*” con los perfiles de los encuestados. Se utilizarán dos tipos de técnicas, descriptivas, sobre todo diagramas de barras y tablas; e inferenciales, contrastes de independencia de Chi-cuadrado (X^2) y técnicas relacionadas como residuos tificados, y V de Cramer.

Análisis descriptivo

El análisis descriptivo es siempre la primera visión de los datos, en este caso se han observado las tablas con las variables explicativas (las cinco variables que se han mencionado anteriormente) cruzadas con la variable explicada (“*compra*”), con el apoyo de diferentes diagramas de barras.

Edad:

Figura 4 Diagrama en frecuencias relativas de los niveles de compra de ropa online en función de la edad

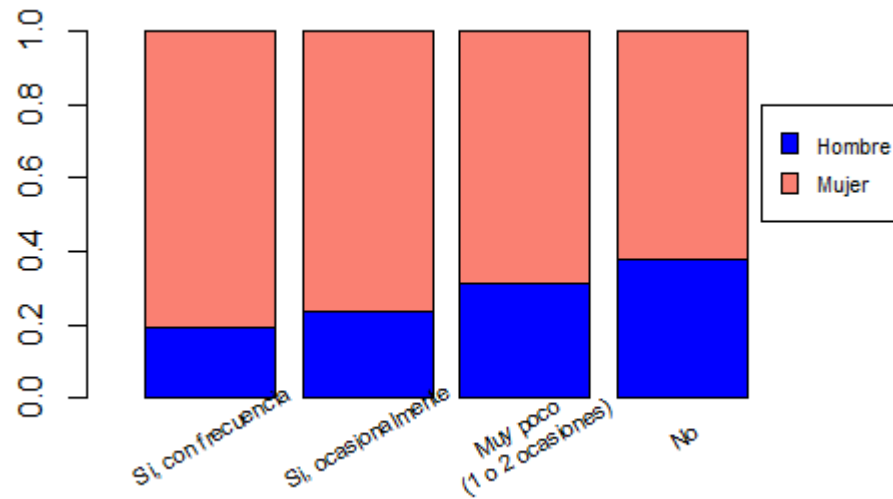


Fuente Elaboración propia en R

En este primer caso, a través del correspondiente diagrama de barras en frecuencias relativas que se puede observar en la **Figura 4** se puede observar claramente como a menor nivel de edad los individuos encuestados tienden a comprar más ropa online. Es así que todos los niveles de “sí compra” son superiores en los tres primeros niveles (“*hasta 25 años*”, “*entre 26 y 35 años*”, “*entre 36 y 45 años*”) mientras que los dos últimos muestran una presencia mayor de “*no compra*” en la web.

Género:

Figura 5 Diagrama en frecuencias relativas de los niveles de compra de ropa online en función del género

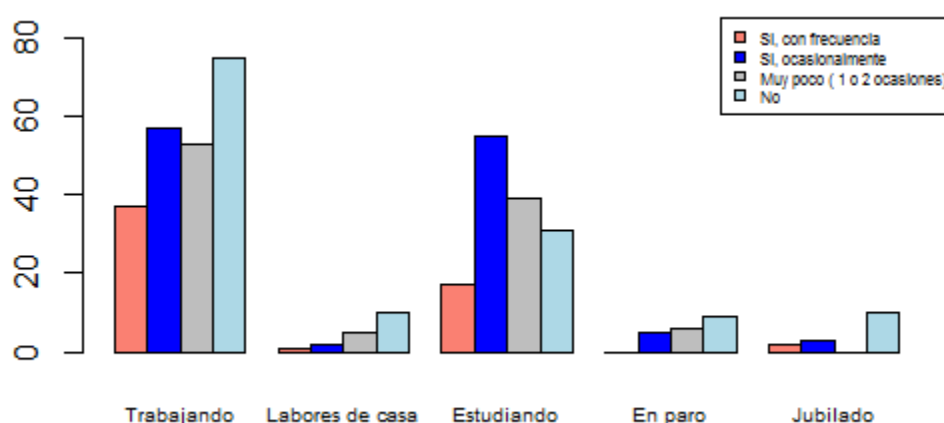


Fuente Elaboración propia en R

En la **Figura 5**, se observa la compra de ropa online en función del género del sujeto encuestado. Además, en ella se aprecia como la proporción de hombres aumenta cuando se reduce la intensidad de compra, de hecho el nivel que mayor representación tiene en los hombres es “no comprar” ropa en la web. Del mismo modo se destaca que la proporción de comprar ropa por internet es mayor en los niveles de “si compra” que de no “compra”, precisamente al contrario que en el género masculino.

Ocupación:

Figura 6 Diagrama en frecuencias absolutas de los niveles de compra de ropa online en función de la ocupación



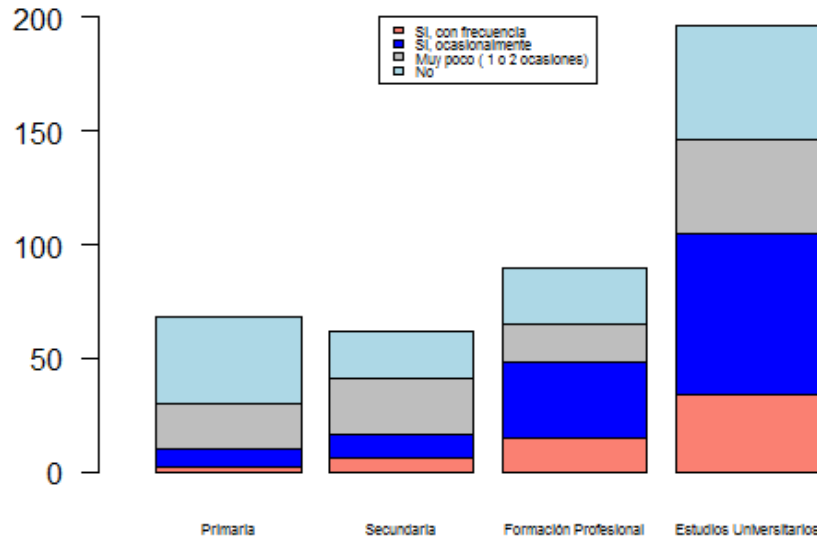
Fuente Elaboración propia en R

En la **Figura 6**, se observa un diagrama de barras en frecuencias absolutas, que representa la variable “*compra*” con respecto a la ocupación de los sujetos encuestados. De esta forma se resaltan que las dos categorías más importantes y relevantes que motivan la compra de ropa online en internet que vienen dadas por los individuos que están “trabajando” o “estudiando”.

Centrándose en esas dos categorías (“trabajando” o “estudiando”), se hacen diferentes observaciones que se citan a continuación.

Por un lado analizando los individuos que no compran ropa por internet: se observa que en el grupo de estudiantes no es la categoría más frecuente, algo que no pasa con los sujetos que estudian ni con los demás tipos de ocupación.

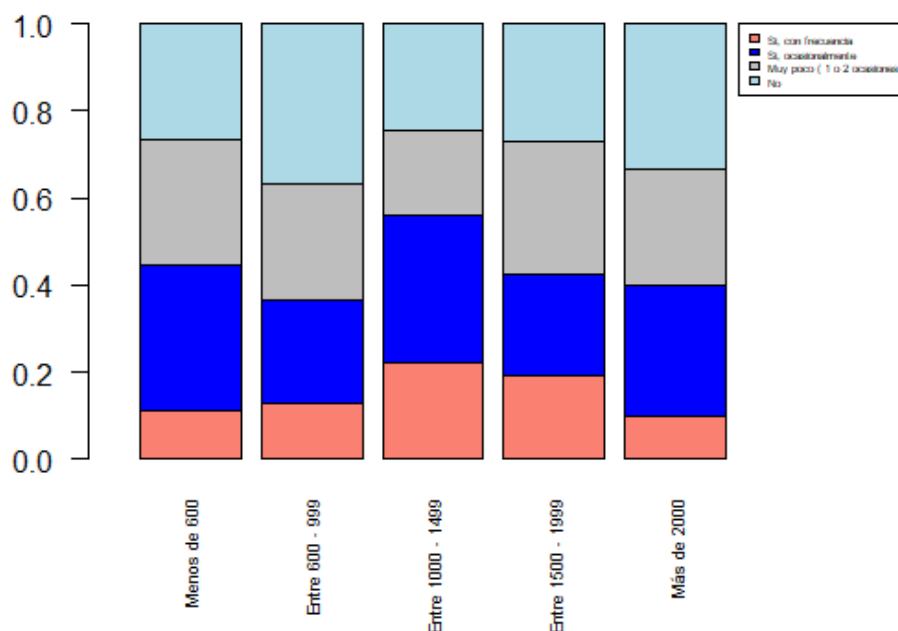
Por otro lado, con respecto a los niveles en los que sí se compra ropa en la web, destacar que, a pesar de que los sujetos que estudian tienen una mayor representación con respecto al conjunto de los niveles de compra, con máximo “sí, ocasionalmente”; son los sujetos que trabajan los que destacan en el nivel comprar “Si, con frecuencia”, el de mayor intensidad de compra.

Estudios:**Figura 7** Diagrama en frecuencias absolutas de los niveles de compra de ropa online con respecto a los estudios

Fuente Elaboración propia en R

Con respecto al diagrama de barras en frecuencias absolutas que se representan en la **Figura 7** se observa la relación bilateral entre los niveles de compra y los estudios de cada individuo. En este caso, se observa como en estudios universitarios y formación profesional el nivel más grande corresponde a comprar “si, ocasionalmente”, y en relación a estos dos niveles de estudio, llama la atención que haya más gente que no compre con “estudios universitarios” que en “formación profesional”. Por otro lado en “estudios primarios” destaca el mayor número de personas que no compran.

En resumen, la pauta es que los individuos con estudios más altos aparecen relacionados con mayor compra de ropa por internet, si bien no es descartable en este caso una influencia de la variable “edad” como se puede observar a través de la **Figura 17** que, “estudios primarios” está asociado a personas más mayores, mientras que estudios universitarios y sobre todo FP se asocia con personas más jóvenes.

Ingresos:**Figura 8** Diagrama en frecuencias relativas de los niveles de compra de ropa online en función de los ingresos individuales del sujeto

Fuente Elaboración propia en R

En base a la **Figura 8**, se puede ver un diagrama de barras en frecuencias relativas de los niveles de compra de ropa online de los encuestados en función de sus ingresos obtenidos. En este caso, se destaca una apreciación muy importante que es la cantidad de individuos que obtienen menos de 600€ mensuales pero sin embargo es una de las categorías de “compra” que cuenta con una mayor proporción. Una de las hipótesis que se plantean al observar esta relación es que se debe a la influencia de otras variables una tercera variable, como es la “edad”, que posiblemente relacione a estos individuos de menores ingresos con una edad más temprana (gente joven o menor de edad que sigue viviendo de sus padres) y sin embargo compran mayor ropa online que otro conjunto (véase **Figura 18**).

Además otra apreciación es que no se cumple el criterio de a mayores ingresos, mayores compra de ropa online. Y esta característica también podría venir definida a raíz de la variable “edad”, la cual podría afectar a que la gente con mayores ingresos sean los que cuenten con mayor edad.

Gracias a la Estadística Descriptiva, que nos permite observar desde un primer momento y de forma más organizada los datos, se concluyen varias relaciones importantes. A modo de resumen y para aclarar las características más destacadas que han surgido:

- Es evidente la primera relación entre compra y edad: cuanto más jóvenes sean los sujetos, mayor es la compra de ropa online.
- Con respecto a la variable género, es evidente que el sexo femenino compra más que el masculino, puesto que las mujeres tienen una mayor proporción respecto de frecuencias relativas obteniendo que son las compradoras de ropa online por excelencia en base al género, en este estudio.
- En base a su ocupación se destaca que en la muestra tienen mayor relevancia los sujetos que “estudian” y “trabajan”. Estando los “estudiantes” entre los mayores compradores, al considerar todos los niveles de compra, pero la gente que trabaja si sólo se considera el nivel de compra más intenso.
- En base a los estudios que tenga el determinado individuo, destacan los sujetos con más estudios (“estudios universitarios” y “formación profesional”), que cuentan con una contribución mayor a la hora de comprar, mientras que “estudios primarios” cuenta con menos. Sin embargo, esta relación puede verse influida por la “edad” como ya se ha mencionado.
- La última variable estudiada en relación con la compra de ropa online son los ingresos obtenidos en los que se destaca la categoría de los sujetos con ingresos inferiores a 600€ los cuales se muestran con una gran participación a la hora de realizar compras de ropa en la Web. Mientras que por otro lado, los individuos que mayores ingresos tienen no se relacionan tanto a la hora de comprar. Esto viene dado por la influencia de la variable “edad” ya comentada.

Una vez determinadas cuales son las características más importantes de la muestra y de las relaciones entre las variables de perfil y la variable “compra” mediante el análisis descriptivo. Ya que el análisis descriptivo no permite afirmar que las diferencias observadas no sean algo más que aleatoriedad. Para poder decir que la diferencia es significativa se debe recurrir a la Estadística Inferencial, que será lo que ocupe la siguiente parte del análisis.

Para ello se buscará determinar si hay relación significativa o no entre las variables explicativas con la variable explicada, con lo cual se aplicarán Test de independencia de Chi-Cuadrado (X^2). Una vez se haya observado si existe o no relación entre las variables, se ha de determinar entre que variables existe más o menos asociación (relación), esto se lleva a cabo mediante el estadístico V de Cramer.

Es esencial determinar desde un primer momento si hay o no relación entre las variables tanto entre las variables explicativas con ellas mismas, como con la variable explicada, ello se realiza a través del Test de independencia de Chi cuadrado (X^2).

Test de independencia con Chi cuadrado (X^2)

Este contraste Chi-Cuadrado, plantea las siguientes hipótesis:

- Hipótesis nula (H_0): ¿los atributos son independientes?
- Hipótesis alternativa (H_1): ¿los atributos son dependientes?

Una vez realizados los cálculos, contrastando independencia entre compra y las cinco variables explicativas, se han obtenido para cada contraste los p-valores que figuran en la **Tabla 7**, en ellos se ve que la única relación que no es significativa es Compra-Ingresos, ya que su p-valor es superior al nivel de significación del 5%. Las demás relaciones se muestran significativas puesto que su p-valor es inferior al 5% de nivel de significación.

Tabla 7 Contrastes Chi-Cuadrado y V de Cramer

Variabes	Chi – Cuadrado (X^2)	V de Cramer
Compra – Edad	2.632e-05	0.1845
Compra – Género	0.02323	0.15083
Compra – Ocupación	0.001281	0.16047
Compra – Estudios	4.332e-07	0.1936
Compra con Ingresos	0.4048	0.1072

Fuente Elaboración propia en R

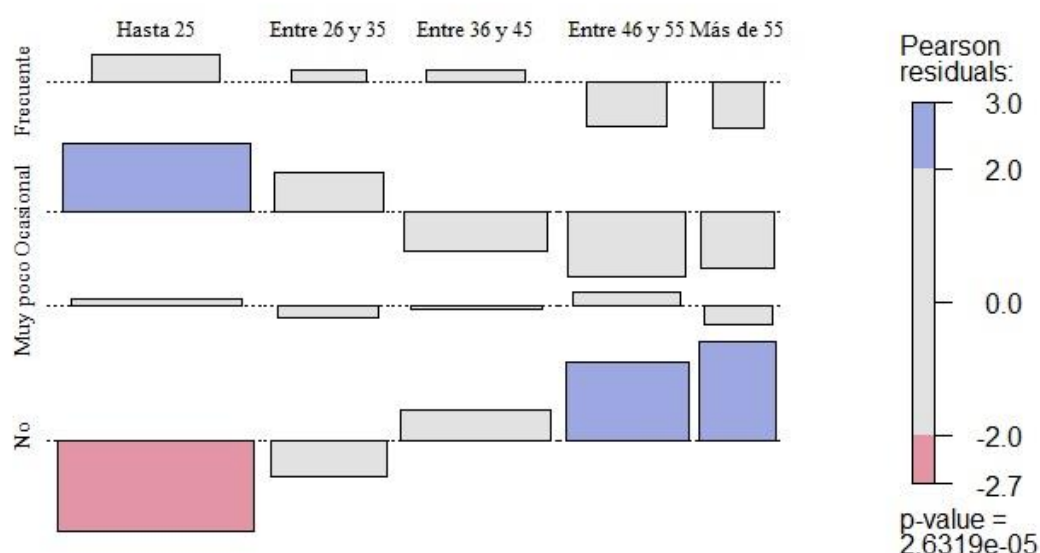
Complementando al contraste X^2 también se analizan sus residuos (las diferencias entre las frecuencia observadas y las frecuencia teóricas). En este caso, se representan los residuos tipificados mediante unos gráficos que se denominan diagramas de

asociación, que se observan a continuación. De esta manera se observará qué categorías son responsables de la falta de independencia y por lo tanto de la relación significativa entre los pares de variables.

En esta parte del análisis no se ha incluido la relación entre Género y Compra puesto que no se obtenían ningún residuo significativo.

Edad:

Figura 9 Diagrama de asociación de la compra de ropa online respecto a la edad

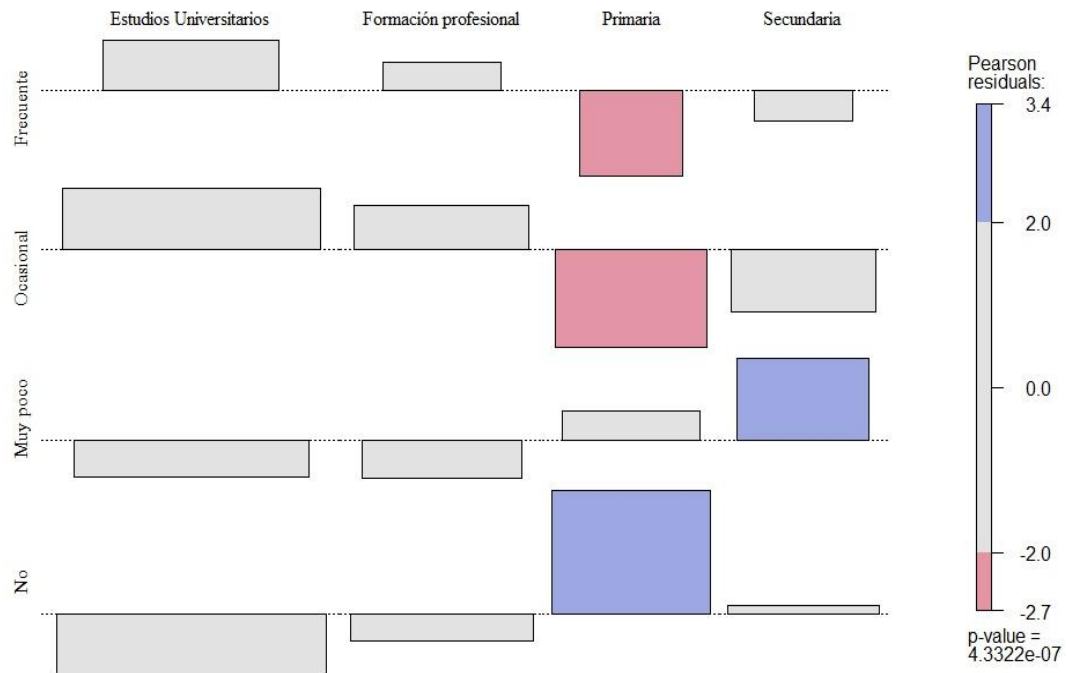


Fuente Elaboración propia en R

En el primer diagrama de asociación (**Figura 9**), los agentes con menos de 25 años están sobrerrepresentados en la compra “si, ocasionalmente” e infrarepresentada en “no compra”, es decir, los jóvenes son un grupo que compran ropa en la web, más que el resto del conjunto de población. Mientras que los otros dos grupos importantes en este caso, que relacionaría la compra con los sujetos entre “46-55 años” y “mayores de 56 años” están sobrerrepresentados en “no comprar” ropa en la web, lo que indica que son los sectores de población menos proclives a la compra de ropa online.

Estudios:

Figura 10 Diagrama de asociación de la compra de ropa online respecto a los estudios

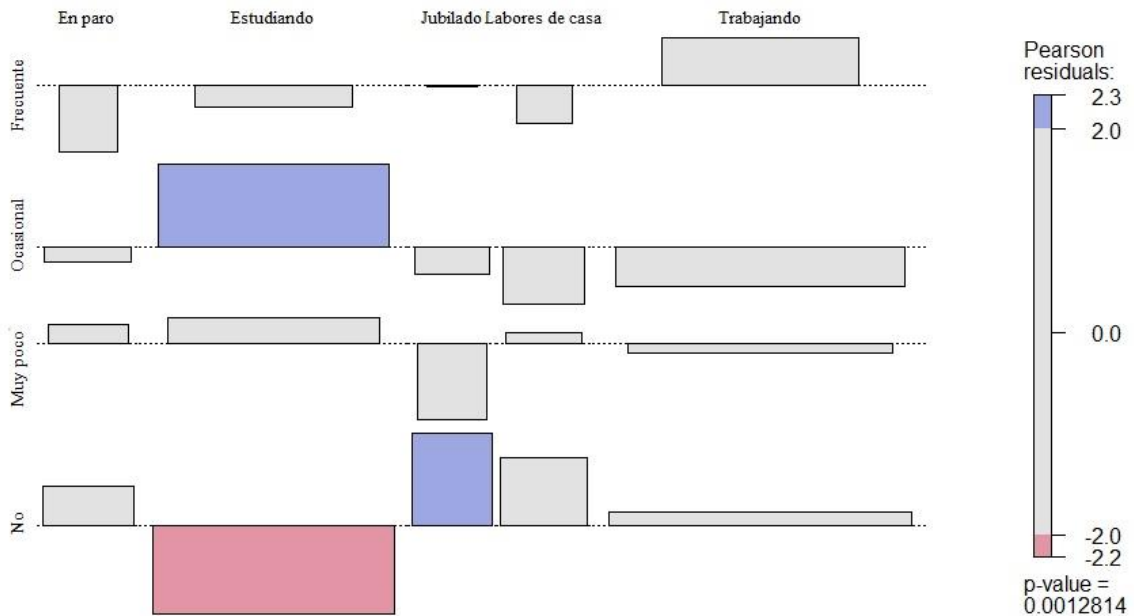


Fuente Elaboración propia en R

En la **Figura 10** se observan los residuos tipificados de la compra online con respecto los estudios del individuo. Es decir, gracias a este gráfico observamos que hay dos grupos que muestran grandes diferencias: en “*primaria*” ya que está infrarepresentada en compra “*si, ocasionalmente*” y “*si, frecuentemente*”, mientras que está sobrerrepresentada en “*no compra*” mientras que en “*secundaria*” está sobrerrepresentada en comprar “*muy poco (1 o dos ocasiones)*”.

Ocupación:

Figura 11 Diagrama de asociación de la compra de ropa online respecto la ocupación



Fuente Elaboración propia en R

Finalmente y como último diagrama de asociación (**Figura 11**) se analiza la relación entre la compra de ropa online y la ocupación de los agentes. En este caso hay dos grupos significativos que marcarían la tendencia de la compra de ropa en la web. El primer grupo viene dado por la compra de ropa por agentes que se encuentran estudiando, en los que se observa que esta variable se encuentra sobre-representada con respecto al nivel de compra “sí, frecuentemente” y al mismo tiempo, infra-representada respecto la ausencia de compra online. Mientras que en el otro grupo se encontrarían los individuos que se encuentran jubilados y que no comprarían en la web.

Una vez determinada la existencia o no de asociación entre las variables, el siguiente paso es medirla. Una de las medidas más usadas es la V de Cramer, que toma valores entre cero (independencia) y 1 (asociación perfecta). Esta medida aunque es un

análisis descriptivo, se coloca en la sección inferencial puesto que se está usando como complemento al contraste de independencia realizado.

Estadístico V de Cramer

En la **Tabla 7**, también se observan los valores correspondientes al Estadístico V de Cramer que reciben las relaciones bidimensionales entre las variables de este estudio. Para este estadístico se observa que cuanto mayor sea su valor mayor será la asociación entre las variables.

Es evidente desde un primer instante que las variables a las que se le da mayor relevancia con respecto a la variable compra son: “*estudios*” (0.194) y “*edad*” (0.185). El resto de variables se enmarcan bajo el siguiente orden de relación: “*ocupación*” (0.160), “*género*” (0.151) y por última y la que cuenta con la menor relación entre ellas sería la variable “*ingresos*” (0.117).

Los resultados hasta aquí obtenidos se refieren a una sola variable o a la combinación de dos de ellas. Sin embargo, estos mismos resultados parecen mostrar una fuerte relación entre variables explicativas, como podría ser el caso de falta de relación significativa entre niveles de Ingreso y de Compra. Por esta razón se completa el análisis con métodos multivariantes, intentando así incluir en nuestro análisis relaciones más completas que las estudiadas en el caso bivalente.

3.2.3 Análisis Multivariante

El desarrollo actual de los métodos de análisis multivariantes (ya comentados anteriormente de forma teórica en el capítulo 2.2) permiten adoptar una aproximación global para el análisis de los datos en las encuestas. Un análisis global, que tenga en cuenta numerosas variables simultáneamente, puede proporcionar información de más calidad y gestionar, en alguna manera, las múltiples relaciones que podrían existir entre las variables explicativas.

En este caso se han elegido dos análisis multivariantes para aplicar de forma práctica en el análisis de la compra de la ropa online: Análisis Cluster y Análisis de Correspondencias Múltiple.

Análisis Cluster

El análisis cluster se define como un método estadístico multivariante de clasificación automática que a partir de una tabla de datos (individuos-variables) trata de situarlos en grupos homogéneos, de manera que los individuos que pueden ser considerados similares sean asignados a un mismo conglomerado (también llamado cluster). En este trabajo el Análisis Cluster se aplica a cuatro variables que son las frecuencias relativas de cada nivel de compra en cada una de las categorías de cada variable explicada, por lo que se manejan 4 variables para 21 individuos (cada una de las categorías consideradas). Su objetivo será clasificar las relaciones que se observaron entre estas categorías en el conjunto de análisis realizados hasta ahora, y de esta manera asociar cada una de ellas a diferentes perfiles de compra que surgirían al realizar este análisis.

En este caso se realiza un cluster jerárquico que se observa en la **Figura 12** a través de un dendograma. Se ha considerado una división de siete conglomerados, teniendo en cuenta los saltos entre las distancias de cada agrupamiento.

Para interpretar esos conglomerados se han calculado los valores medios de las frecuencias de compra en cada uno de ellos, como figura en la **Tabla 8**. Basándonos en esos valores se observa que seguir la frecuencia de los que no compran facilita la descripción de cada grupo. Teniendo eso en cuenta y con la ayuda del dendograma (**Figura 12**) hemos sacado una serie de conclusiones en base a los perfiles de los consumidores de ropa en la web, entre las que se distinguen las siguientes:

Desde la perspectiva de usuarios que menos compras de ropa online realizan (**véase Tabla 8**) se observan cuatro primeros conglomerados (grupo seis, tres, cuatro y siete) con una alta frecuencia de “no compra”.

El grupo seis, que incluye a los individuos que están jubilados, cuentan con la mayor participación de no compradores respecto no realizar la compra de ropa online

(exactamente un 66%), es decir, los jubilados forman el grupos de agentes que menos ropa compra mediante la web.

El siguiente grupo con mayor porcentaje de “*no compra*” es el grupo tres, aquí se sitúan los agentes que cuentan con edades más próximas a ser jubilados. Además se integran los sujetos que cuentan con estudios de Primaria-EGB-ESO; esta categoría, bajo estos datos, está asociada a una alta edad del individuo, lo que refuerza la idea de que los individuos con mayor edad son los que intervienen menos en la compra de ropa online. También se incluirían ahí los agentes que realizan labores de casa.

En el mismo sentido, otro grupo que tampoco interviene en gran medida en la compra de ropa online es el grupo cuatro. En él se encuentran los individuos que están en paro, con una frecuencia de “*no compra*” del 45%.

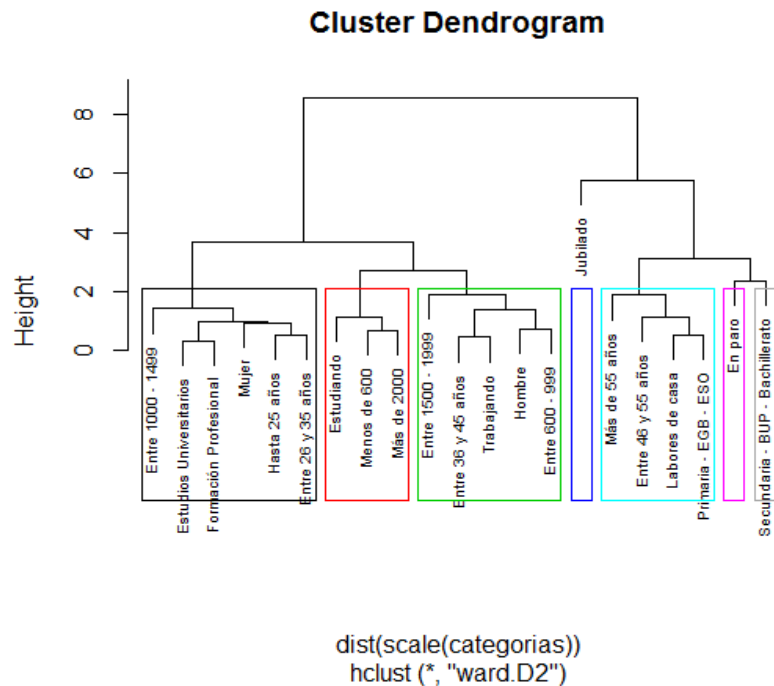
Finalmente, el cuarto grupo entre los de más “*no compra*” es el cuatro, formado por los individuos con estudios de Secundaria-BUP-Bachillerato, aun así, de entre los cuatro grupos con “*no compra*” es el que menos participación tiene, con un 33%. Respecto de este grupo, a pesar de que se puede ver que en el dendograma su agrupamiento en los clusters de “*no compra*”, no lo es por su frecuencia en esta categoría (superada por el grupo 1, con un 35%) si no por el efecto asociado con la otra categoría de menos compra, la de “Muy poco”, con una frecuencia del 38% (el grupo 1 tiene un 26%).

Por otro lado, se sitúan los tres grupos restantes de usuarios. En este caso continuando con la misma perspectiva del grupo que menos compra entre estos conglomerados, se sitúa el grupo uno. En él se identifican las categorías de “*hombre*”, “*entre 36 y 45 años*”, “*trabajando*”, “*ingresos entre 600€ y 999€*” e “*ingresos entre 1500€ y 1999€*”. Aunque cuenta con una gran participación en no realizar compras por internet o realizarlas muy poco. Se observa que es uno de los conglomerados que también cuenta con una gran participación en comprar frecuentemente (19%).

Finalmente se sitúan los dos grupos con una menor participación respecto no comprar por internet y por el contrario cuentan con gran participación a la hora de sí realizarla. Tanto el perfil del grupo dos (en los que se incluyen las categorías de “*Mujer*”, “*Entre 25 y 35 años*”, “*Estudios Universitarios*”, “*Formación Profesional*” e “*Ingresos*

entre 1000€ y 1499€”, como el perfil del grupo cinco (“Estudiando”, “Menos de 600€” y “Más de 2000€”) estarían relacionados en la existencia de una mayor compra de ropa online, diferenciándose en que el grupo 2 es más comprador (categoría “si, frecuentemente” en un 17%) que el grupo 5 (categoría “si, frecuentemente” 10%).

Figura 12 Dendrograma



Fuente Elaboración propia en R

Sintetizando, se obtiene que en los 3 clusters que más compran vienen integrados por los individuos de menor edad (hasta 45 años) y mayor nivel de estudios. Es interesante observar como el agrupamiento de los niveles de ingreso, no tiene una relación lineal con los niveles de compra.

Tabla 8 Distribuciones de la compra de ropa online con respecto a sus variables explicativas

	Si, con frecuencia	Si, ocasionalmente	Muy poco	No	Grupos
Hombre	0,08943089	0,2357724	0,2601626	0,4146341	1
Entre 36 y 45 años	0,15151515	0,2323232	0,2424242	0,3737374	1
Trabajando	0,16666667	0,2567568	0,2387387	0,3378378	1
Entre 600 - 999	0,12658228	0,2405063	0,2658228	0,3670886	1
Entre 1500 - 1999	0,19230769	0,2307692	0,3076923	0,2692308	1
Media grupo 1	0,145300536	0,23922558	0,26296812	0,35250574	
Mujer	0,1559322	0,3186441	0,240678	0,2847458	2
Hasta 25 años	0,16071429	0,3809524	0,2559524	0,202381	2
Entre 26 y 35 años	0,15517241	0,3793103	0,2241379	0,2413793	2
Estudios Universitarios	0,17346939	0,3622449	0,2091837	0,255102	2
Formación Profesional	0,16666667	0,3666667	0,1888889	0,2777778	2
Entre 1000 - 1499	0,2195122	0,3414634	0,195122	0,2439024	2
Media grupo 2	0,171911193	0,358213633	0,21899382	0,25088138	
Entre 46 y 55 años	0,07575758	0,1666667	0,2727273	0,4848485	3
Más de 55 años	0,03846154	0,1153846	0,1923077	0,6538462	3
Labores de casa	0,05555556	0,1111111	0,2777778	0,5555556	3
Primaria - EGB - ESO	0,02941176	0,1176471	0,2941176	0,5588235	3
Media grupo 3	0,04979661	0,127702375	0,2592326	0,56326845	
En paro	0	0,25	0,3	0,45	4
Estudiando	0,11971831	0,3873239	0,2746479	0,2183099	5
Menos de 600	0,10958904	0,3356164	0,2876712	0,2671233	5
Más de 2000	0,1	0,3	0,2666667	0,3333333	5
Media grupo 5	0,109769117	0,3409801	0,2763286	0,27292217	
Jubilado	0,13333333	0,2	0	0,6666667	6
Secundaria - BUP - Bachillerato	0,09677419	0,1774194	0,3870968	0,3387097	7

Fuente Elaboración propia en R y Excel

A lo largo de este análisis se han hecho varias aclaraciones en base a los otros métodos llevados a cabo anteriormente para explicar las relaciones de la variable compra con sus variables explicativas.

Se ha establecido que respecto la dependencia y significación las dos variables más importantes que existe con relación a la compra online venía dada por la edad y los estudios.

Esto tampoco determinada exactamente ninguna categoría explicase mejor que otra, con lo que se ha llevado a cabo el Análisis Cluster en el que nos aparecen una

serie de individuos con sus distinguidos perfiles de compra pero tampoco se llega a comparar explícitamente todas las variables al mismo momento.

Análisis de correspondencias múltiples

El Análisis de Correspondencias Múltiple (ACM) es un método de análisis multivariante que trata la relación entre múltiples variables, reorganizándola y creando nuevas variables, independientes entre sí pero que se pueden dar un nuevo sentido a las diferentes relaciones que recogen.

Aplicándolo se espera ver no solo las relaciones entre “compra” y las demás variables explicativas sino también entre el conjunto de todas las categorías.

Bajo la **Figura 13** se observa un mapa perceptual en el que se representan la variable explicada (compra)³ junto con sus variables explicativas (edad, sexo, ocupación, estudios e ingresos) pero en este caso bajo dos dimensiones, creadas por el análisis. Es interesante observar cuál de ellas se asocia más con la variación de “compra”, por ello se averigua observando sus contribuciones en las dimensiones en la **Tabla 9** y éstas obtienen una serie de valores nuevos para cada variable. También se percibe a través de la tabla que estas dimensiones obtienen una serie de valores nuevos para cada variable. Al observar estas contribuciones se puede ver que los valores más altos de la variable vienen dadas por la dimensión dos (exactamente las categorías extremas “No” y “Si, con frecuencia”).

Además otra de las claves a destacar en este caso, es que en la dimensión dos donde la variable “compra” tiene un orden lógico en sus categorías, en primer lugar se observa “Si con frecuencia”, a continuación “Si, ocasionalmente”, después “Muy poco (1 o 2 ocasiones)” y finalmente “No comprar”.

Al centrarse en esta dimensión se observa, a la altura de “Si, con frecuencia” tres categorías: “Formación profesional”, “Entre 26 y 35 años” y “Entre 1000 -1499”, lo que formaría, según este método, el perfil del individuo más comprador.

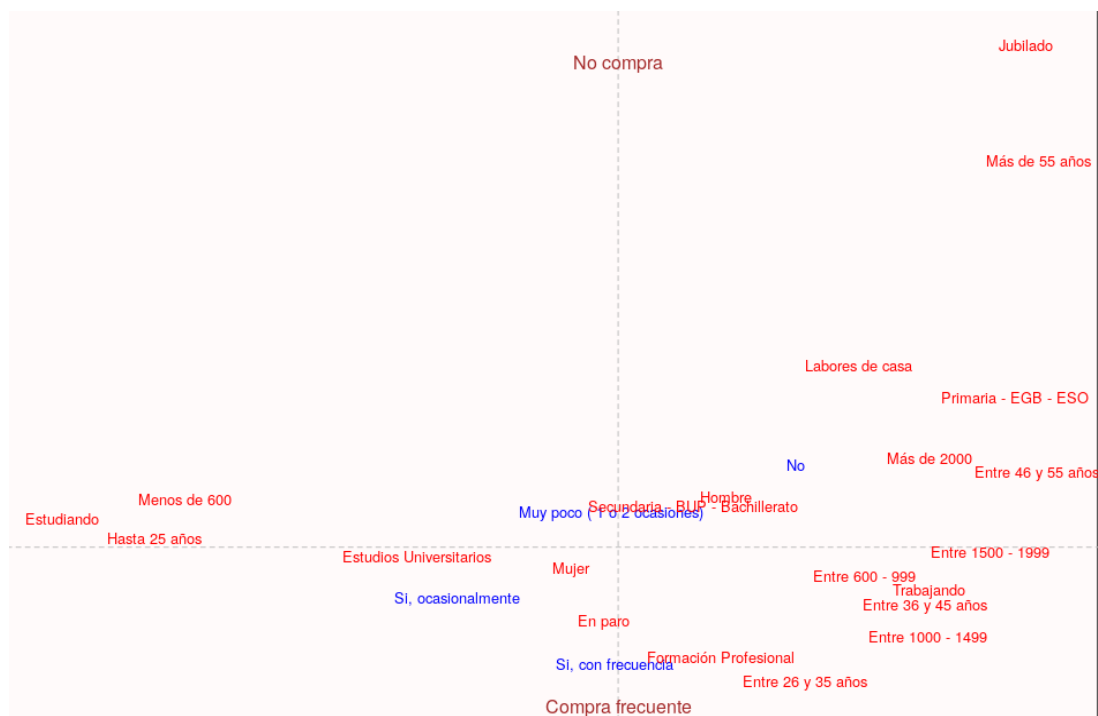
³ Se destaca que el análisis de correspondencias no separa entre variables explicadas y explicativas, pero dado que nuestro interés sigue siendo explicar las categorías de compra en función de las otras variables seguiremos manteniendo esa distinción.

Se observa también que los sujetos que están jubilados están alejados en la posición de “no compra”, es decir, no muestran una relación muy intensa con respecto a las otras categorías y con ello muestran más diferencias a la hora de realizar compras por internet.

Menos distanciado pero también en situación de “no compra” estarían “labores de casa” y “Primaria-EGB-ESO”.

Una posición sorprendente es la categoría “en paro” ya que se muestra en una posición de compra relativamente alta, sin embargo, observando la **tabla 9** se ve que tiene una contribución baja, por lo que no se toma como un resultado relevante.

Figura 13 Mapa perceptual de las dimensiones 1 y 2



Fuente Elaboración propia en R

Tabla 9 Valores de las variables sobre las cinco dimensiones ordenados bajo la Dim 2

	Dim1	Dim2
Entre 1500 - 1999	1,94	0
Hasta 25 años	17,09	0,08
Estudios Uni.	3,73	0,11
Entre 600 - 999	2,59	0,52
En paro	0	0,72
Secundaria - BUP -Bac	0,17	0,72
Estudiando	20,5	0,73
Muy poco (1 o 2)	0	0,78
Mujer	0,3	0,94
Más de 2000	1,56	1,79
Hombre	0,69	2,13
Si, ocasionalmente	1,55	2,17
Entre 36 y 45 años	4,21	2,19
Entre 46 y 55 años	5,48	2,35
Menos de 600	14,77	2,45
Labores de casa	0,37	2,96
Trabajando	10,44	2,96
Entre 1000 - 1499	4,17	4,74
No	1,79	5,37
Si, con frecuencia	0	5,39
Formación Profesional	0,46	7,23
Entre 26 y 35 años	1,06	7,64
Primaria - EGB - ESO	4,6	9,05
Jubilado	0,8	16,93
Más de 55 años	1,71	20,05

Fuente Elaboración propia en R y Excel

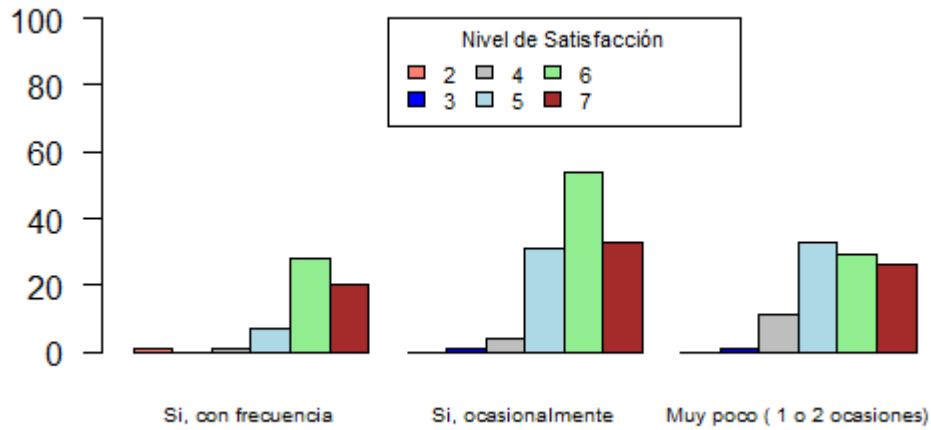
3.3 Estudio sobre otros aspectos

En este apartado, a mayores de haber realizado el análisis estadístico con mayor detenimiento en las cinco variables anteriores, y puesto que en el cuestionario también se han llevado a cabo otras preguntas, se ha hecho dos breves análisis sobre el nivel de de satisfacción que tiene el cliente en la compra de ropa online. Y, para los que “no compran”, las explicaciones (motivos) más relevantes de que no lo hagan.

De este modo y recurriendo de nuevo al análisis bivariante descriptivo, a través de la **Figura 13 y 14**, se describe el grado de satisfacción que tiene el cliente al realizar la compra online. Se pueden apreciar que los agentes cuentan con mayor satisfacción, cuantas mayores compras de ropa online realizan. Por lo general los usuarios que compran ropa por internet tienen buena valoración de ello, ya que en muy pocos la valoración queda por debajo del nivel de satisfacción 5. Además, a pesar de que la

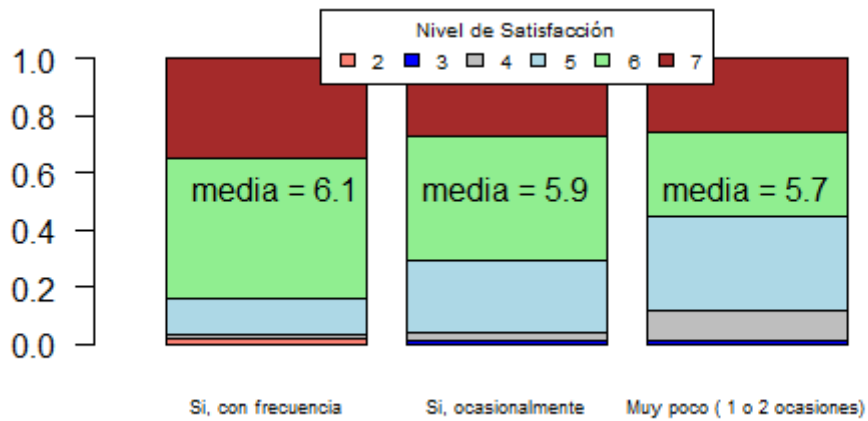
media no tiene grandes diferencias entre niveles, se puede observar cómo los valores de satisfacción son más altos a medida que aumenta la frecuencia de compra.

Figura 14 Grado de satisfacción del cliente en frecuencias absolutas



Fuente Elaboración propia en R

Figura 15 Grado de satisfacción del cliente en frecuencias relativas

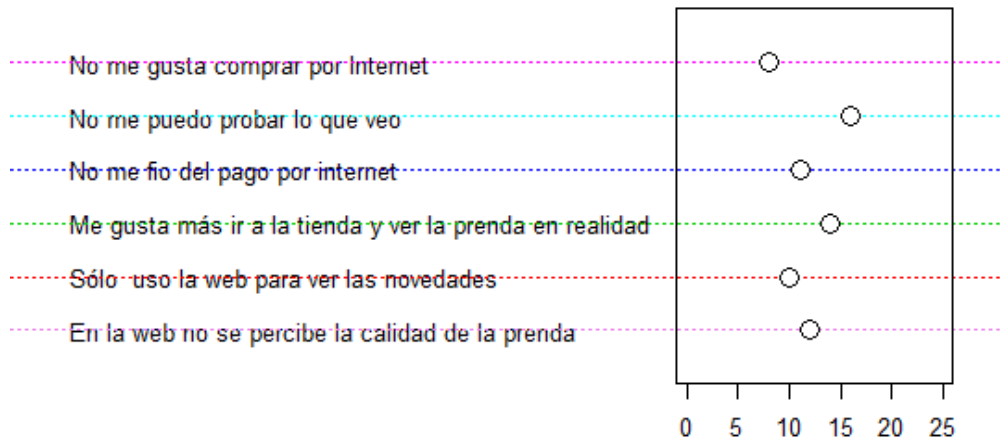


Fuente Elaboración propia en R

Por otro lado, también se destaca a través de la **Figura 17** los motivos que más influyen a la hora de que los individuos no compren ropa online, indicando en la gráfica el número de veces que se menciona cada causa. Uno de los motivos que más influencia tiene es que los individuos no pueden probar lo que ven antes de realizar la compra. Por otro lado también se destaca que realmente a los clientes les gusta más asistir a la

tienda física y poder observar la prenda en realidad. De este modo se destaca que a pesar de que el comercio de ropa online se ha incrementado continúan las preferencias de observar el producto, poderlo tocar, sentir... Aunque los otros motivos también están presentes en cantidades cercanas a estas dos más citadas.

Figura 16 Diagrama de puntos de los usuarios que no compran en relación con sus motivos



Fuente Elaboración propia en R

Conclusiones

Uno de los objetivos de este estudio era mostrar el uso de la Estadística en el ámbito de la Economía más concretamente, en la Investigación de Mercados. Para ello se ha realizado en el primer capítulo una descripción de varias técnicas estadísticas aplicadas en esa área.

Como parte de ese objetivo se ha realizado un estudio del perfil de los consumidores de moda online, abarcando desde el diseño del cuestionario a los diferentes tipos de técnicas descritas en el primer capítulo de este trabajo.

Desde el primer instante se confirma que hay un sesgo en la muestra a raíz de recoger la información mediante un cuestionario, lo que genera que la muestra no sea aleatoria. Los datos los hay que interpretar con cautela puesto que aunque se extraigan perfiles de compra de los consumidores de ropa online, no existen garantías de que se puedan extrapolar a la totalidad de la población.

El objetivo de este análisis ha sido establecer unos métodos estadísticos a una muestra que estudia el comercio de ropa online y mediante ella dar respuesta a las siguientes preguntas: ¿Quiénes compran más? .Y lo más importante, ¿Por qué compran más?, ¿Hay alguna característica en común entre los sujetos que motive esa compra online?

Estas preguntas clave, ya realizadas a lo largo del trabajo, se responden a través de los datos extraídos y cuyas conclusiones generan unos determinados perfiles de compra de los consumidores de ropa online.

Haciendo una síntesis de los resultados logrados se pueden destacar varias conclusiones:

Los factores más determinantes a la hora de realizar la compra de ropa online son dos variables: edad y estudios. Además estas dos variables interfieren a las otras variables en gran medida, como ya se ha observado a lo largo del análisis, y que tienen como consecuencia principal que el nivel de ingresos no aparezca como determinante para realizar la compra online.

Además de estas conclusiones generales se han podido definir tres perfiles claros en los sujetos que compran o no ropa online, como son:

El perfil que destaca como mayor comprador potencial viene determinado por clientes que cumplen las siguientes características: “*Formación Profesional*”, “*26 y 35 años*”, “*trabajando*” (aunque no aparezca dentro del perfil se deduce porque obtiene ingresos) y “*Entre 1000€ y 1499€*”, en este caso se observa que es gente con edad más temprana, es decir, joven. Y que además tienen un nivel de estudios alto (aunque no sea el máximo), concordando con el hecho de que edad y nivel de estudios son las variables más influyentes. De este modo, se tiene como mayores compradores a sujetos que se encuentran trabajando y obteniendo un ingreso de entre 1000€ y 1500€, es totalmente razonable que estos clientes puedan ser los que más contribuyen al comercio de ropa online.

Como segundo perfil, nivel de comprador intermedio, se encontrarían las siguientes categorías: “*estudiando*”, “*menos de 600€*” y “*hasta 25 años*”. También aparece aquí la relación existente con la edad, y de forma indirecta con el nivel de estudios. En este caso es gente más joven, más proclive a la compra por internet, con un nivel de formación que se puede suponer importante (siguen estudiando) pero con pocos ingresos. Se destaca aquí la influencia que la edad puede tener sobre los ingresos, ya que a lo largo del trabajo estos no siguen una tendencia lineal con respecto a la intensidad de compra (es decir, no se confirma que cuanto mayores ingresos se obtengan, mayores serán las compras de ropa online), aquí surge la idea de que posiblemente, aunque este segundo perfil compre moda online ocasionalmente, lo hará gracias a la unidad familiar.

El último perfil viene integrado por “*jubilados*”, “*labores de casa*” y “*primaria*”, este grupo destaca por ser el que menos compras de ropa online realiza. Es por ello que aquí se confirma de nuevo la relación anteriormente descrita por las variables edad y estudios; es decir; los sujetos que cuentan con menores estudios, que posiblemente sean los de más edad compran menos ropa por este medio.

Bibliografía

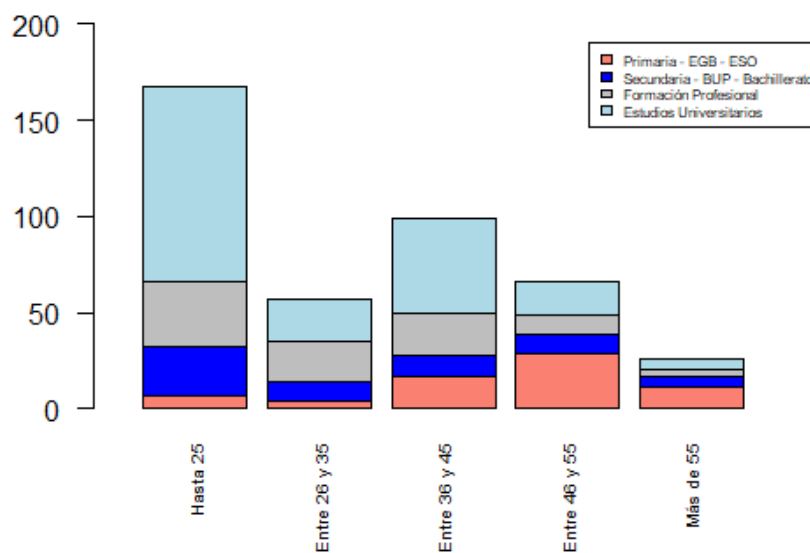
- American Marketing Association. (2013). *Definition of Marketing*. Recuperado de <https://www.ama.org/AboutAMA/Pages/Definition-of-Marketing.aspx>
- Abascal E. y Grande I., (2005). *Análisis de encuestas*. Madrid: ESIC
- Clausen, S. E. (1998). *Applied Correspondence Analysis: An Introduction*. Thousand Oaks, CA : Sage
- Escuder R.y Santiago J., (1995). *Estadística aplicada. Economía y ciencias sociales*. Valencia: Tirant Lo Blanch.
- Fuente F. de la. (2011). *Análisis de Correspondencias Simples y Múltiples*. Recuperado de <https://goo.gl/YJ4Wd5>
- Ihaka, R. & Gentleman, R. (1997). *The R Project for Statistical Computing*. Recuperado de <https://www.r-project.org/>
- Malhotra, Naresh K., (2008). *Investigación de mercados (5ª Ed.)*. México: Pearson Educación
- Marqués, Felicidad (2015). *Técnicas de Investigación de Mercados*. Madrid: RC Libros.
- Martín J. P., Lafuente, M. y Faura, U. (2015). *Guía Práctica de Estadística aplicada a la empresa y al marketing*. Madrid: Paraninfo.
- Martín Pliego, Fco. J. (2004). *Introducción a la Estadística Económica y Empresarial Teoría y Práctica (3ª Ed.)*. Madrid: Thomson.
- Olmo Arriaga, José Luis, del y Fondevila, Joan Francesc. (2009). La comunicación como herramienta estratégica en la industria de la moda. *Revista de la SEECI*, (18), 1-30. doi: <http://dx.doi.org/10.15198/seeci.2009.18.1-30>
- Organización Mundial del Comercio. (1998). *Programa de trabajo sobre el comercio electrónico (WT/L/274)*. Recuperado de https://www.wto.org/spanish/tratop_s/ecom_s/ecom_s.htm

- Pérez López, C. (2007). *Minería de Datos. Técnicas y Herramientas*. Madrid: Thomson.
- Romero Aroca, P., Lázaro García, C. y González López, J.J. (2013). Estadística descriptiva e inferencial. En Pedro Beneyto Martín, *De la idea a la publicación científica manual de investigación clínica* (pp.166-176). Madrid: Sociedad Española de Oftalmología.
- Ruiz-Maya, Luis (2000). *Métodos estadísticos de investigación de las Ciencias Sociales: técnicas no paramétricas*. Madrid: AC.
- Stevens S.S. (1946). Sobre la teoría de escalas de medición. *Science*, 103(2684), 677-680. Recuperado de <https://goo.gl/tU6n8S>
- Trespalacios J.A., Vázquez, R. y Bello, L., (2005). *Investigación de Mercados: Métodos de recogida y análisis de la información para la toma de decisiones en Márketing*. Madrid: Thomson
- Vargas Sabadía, A. (1995). *Buscando sobre la Estadística descriptiva y Estadística inferencial*. Cuenca: Universidad de Castilla la Mancha. Recuperado de <https://goo.gl/f7WV7a>
- Vidal R. de la. (2011). *Ventajas e inconvenientes de la encuesta por Internet*. Universidad Pública de Navarra
- William G. y Barry J., (2009). *Investigación de mercados*. (9ª Edición). México: Cengage Learning

Anexos

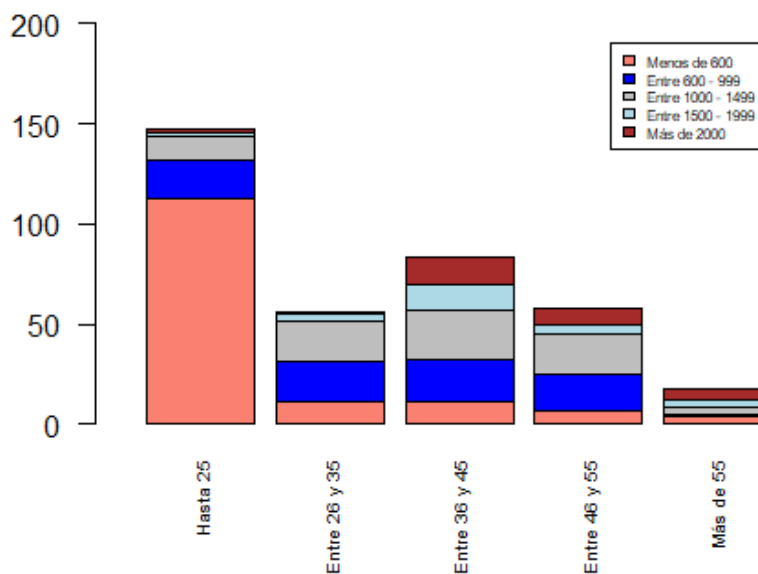
Anexo A Gráficos

Figura 17 Diagrama en frecuencias absolutas de los estudios en función de la edad



Fuente Elaboración propia en R

Figura 18 Diagrama en frecuencias absolutas de los ingresos en función de la edad



Fuente Elaboración propia en R

Anexo B Tablas

Tabla 10 Coordenadas de las dimensiones

	1 Dim	2 Dim
Entre 26 y 35 años	0,43	-0,92
Si, con frecuencia	-0,01	-0,8
Formación Profesional	0,23	-0,75
Entre 1000 - 1499	0,7	-0,6
En paro	-0,03	-0,51
Entre 36 y 45 años	0,7	-0,41
Si, ocasionalmente	-0,37	-0,35
Trabajando	0,71	-0,3
Entre 600 - 999	0,57	-0,21
Mujer	-0,11	-0,15
Estudios Universitarios	-0,46	-0,06
Entre 1500 - 1999	0,84	-0,03
Hasta 25 años	-1,06	0,06
Estudiando	-1,26	0,19
Muy poco (1 o 2 ocasiones)	-0,02	0,23
Secundaria - BUP - Bac	0,17	0,28
Menos de 600	-0,98	0,32
Hombre	0,25	0,35
Entre 46 y 55 años	0,95	0,5
No	0,41	0,57
Más de 2000	0,71	0,61
Primaria - EGB - ESO	0,9	1,02
Labores de casa	0,55	1,24
Más de 55 años	0,96	2,64
Jubilado	0,92	3,43

Fuente Elaboración propia en R y Excel

Tabla 11 Valores del contraste V de Cramer

	Genero	Edad	Ocupacion	Estudios	Ingresos	Compra
Genero	1.000	0.118	0.208	0.069	0.210	0.151
Edad	0.118	1.000	0.481	0.265	0.347	0.185
Ocupacion	0.208	0.481	1.000	0.268	0.392	0.160
Estudios	0.069	0.265	0.268	1.000	0.173	0.194
Ingresos	0.210	0.347	0.392	0.173	1.000	0.117
Compra	0.151	0.185	0.160	0.194	0.117	1.000

Fuente Elaboración propia en R

Tabla 12 P-valores del contraste de Chi-Cuadrado (X^2)

	Genero	Edad	Ocupacion	Estudios	Ingresos	Compra
Genero	0.000000	0.212616	0.001215	0.580975	0.006708	0.023231
Edad	0.212616	0.000000	0.000000	0.000000	0.000000	0.000026
Ocupacion	0.001215	0.000000	0.000000	0.000000	0.000000	0.001281
Estudios	0.580975	0.000000	0.000000	0.000000	0.005414	0.000000
Ingresos	0.006708	0.000000	0.000000	0.005414	0.000000	0.453599
Compra	0.023231	0.000026	0.001281	0.000000	0.453599	0.000000

Fuente Elaboración propia en R

Anexo C Cuestionario

1. Género

- Hombre
- Mujer

2. Edad

- Hasta 25 años
- Entre 26 y 35 años
- Entre 36 y 45 años
- Entre 46 y 55 años
- Más de 55 años

3. Ocupación

- Estudiando
- Trabajando
- Labores de casa
- En paro
- Jubilado

4. Nivel de estudios

- Primaria - EGB - ESO
- Secundaria - BUP - Bachillerato
- Formación Profesional
- Estudios Universitarios

5. Su nivel de ingresos mensuales

Elige ▼

6. ¿Compra habitualmente ropa por internet?

- Sí, con frecuencia
- Sí, ocasionalmente
- Muy poco (1 o 2 ocasiones)
- No

Instrucciones

Instrucciones

-Si su respuesta a la pregunta anterior N°6 es NO pase a la pregunta N°10.
-En caso contrario, continúe en la siguiente pregunta.

7.¿En qué web ha comprado ropa? (puede indicar más de una opción si lo desea)

- Inditex (Zara, Pull&Bear, Massimo Dutti, Bershka, Stradivarius, Oysho, Zara Home y Uterqüe)
- El Corte Inglés
- Privalia
- Mango
- Amazon
- Zalando

Otras (especifique)

8.Motivos por los que compra ropa online (puede indicar más de una opción si lo desea)

- Me resulta más comodo, no me tengo que desplazar
- No tengo que hacer colas
- Encuentro más variedad que en la tienda física
- Hay más promociones y ofertas

Otros motivos (especifique)

Tu respuesta _____

9.Su grado de satisfacción en la compra de ropa online

	1	2	3	4	5	6	7	
1=Totalmente insatisfecho	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	7= Totalmente satisfecho

Sólo en el caso de que su respuesta a la pregunta N° 6 haya sido NO, debe contestar las siguientes preguntas:

10. ¿Motivo por el que no compra ropa online? (puede indicar más de una opción si lo desea)

- No me gusta comprar por Internet
- Me gusta más ir a la tienda y ver la prenda en realidad
- No me fío del pago por internet
- No me puedo probar lo que veo
- En la web no se percibe la calidad de la prenda
- Sólo uso la web para ver las novedades

Otros motivos (especifique)

Tu respuesta _____

11. ¿Ve probable realizar en algún momento futuro su compra online?

- Si
- No