



UNIVERSIDAD DE SALAMANCA

GRADO EN ESTADÍSTICA
FACULTAD DE CIENCIAS

Trabajo de Fin de Grado

**ESTUDIO DEL COMPORTAMIENTO DEL CONSUMIDOR DURANTE
LA COVID-19 EN LOS SUPERMERCADOS: UNA PERSPECTIVA
MULTIVARIANTE**

STUDY OF CONSUMER BEHAVING DURING COVID-19 IN THE
SUPERMARKES: A MULTIVARIAN PERSPECTIVE

Autor: Clara Palomar Sevillano

Tutora: Dra. María Carmen Patino Alonso

UNIVERSIDAD DE SALAMANCA

**GRADO EN ESTADÍSTICA
FACULTAD DE CIENCIAS**

Trabajo de Fin de Grado

**ESTUDIO DEL COMPORTAMIENTO DEL CONSUMIDOR DURANTE
LA COVID-19 EN LOS SUPERMERCADOS: UNA PERSPECTIVA
MULTIVARIANTE**

**STUDY OF CONSUMER BEHAVING DURING COVID-19 IN THE
SUPERMARKES: A MULTIVARIAN PERSPECTIVE**

Tutora: Dra. María Carmen Patino Alonso

Autor: Clara Palomar Sevillano

A handwritten signature in blue ink, appearing to read 'Clara', with a large, sweeping underline that loops back under the name.

ÍNDICE

1. INTRODUCCIÓN	4
2. MARCO TEÓRICO	5
3. OBJETIVOS	8
4. METODOLOGÍA	9
4.1 DESCRIPCIÓN DE LOS DATOS	9
4.2 ANÁLISIS ESTADÍSTICOS	9
4.2.1 FIABILIDAD	10
4.2.2 ANÁLISIS FACTORIAL (AF)	10
4.2.3 ANÁLISIS DE CORRESPONDENCIAS (ACC)	29
5. RESULTADOS	34
5.1 ANÁLISIS DESCRIPTIVO DE LOS DATOS	34
5.2. ESTRUCTURA FACTORIAL DEL CUESTIONARIO	35
5.3. PERFIL DE LOS CONSUMIDORES DURANTE LA COVID-19.....	39
6. CONCLUSIONES	43
7. BIBLIOGRAFÍA	44
8. ABSTRACT	46

1. INTRODUCCIÓN

A principios del año 2020 llegó a España un virus originado en Wuhan (China) a finales de 2019, llamado SARS-CoV-2 que hasta julio de 2021 ha provocado la muerte de más de 80 mil personas en España. Es un virus muy contagioso que produce síntomas similares a los de la gripe y en casos graves produce neumonía y dificultad respiratoria. En pocos meses se empezó a vivir una situación que quedará marcada como una de las pandemias mundiales más grandes de los últimos 100 años.

El día 11 de marzo de 2020 la OMS declaró una pandemia global producida por el COVID-19 y esta pandemia ha provocado muchos cambios en la vida de los habitantes españoles. En primer lugar, el gobierno español decretó un estado de alarma que prohibía la salida de los domicilios excepto para realizar trabajos esenciales y para comprar productos de primera necesidad. Esta crisis sanitaria dio lugar a una gran crisis económica producida por el cierre de miles de negocios, infinidad de españoles mandados a casa con ERTE por falta de trabajo, ser personas de riesgo o haberse contagiado del virus.

Tanto la crisis como el miedo al contagio y las restricciones del gobierno hicieron que muchas familias tuvieran que cambiar sus hábitos de consumo. Menos compras y más grandes, para llenar la despensa y evitar ir al supermercado numerosas veces. Hubo quién cambió su lugar habitual de compra por tener que ir al más cercano o por apoyar a los negocios locales.

Es interesante e importante conocer como ha influido la enfermedad Covid-19 en la vida de las personas. Ha producido giros radicales en la vida de muchos individuos, millones de contagios, miles de muertes, cambios en la economía de los núcleos familiares, distanciamientos sociales, pérdidas de cursos académicos...

A dieciséis meses desde el comienzo de la pandemia y con el 55% de la población española vacunada aun no sé conoce como avanzará cada mes la situación y mucho menos como y cuándo acabará. Lo más predecible es que muchos de los cambios que ha provocado la pandemia se adopten ya como rutinarios tras su finalización.

El objetivo de este trabajo es conocer cómo se comportan los consumidores españoles en los supermercados tras el comienzo de la pandemia desde una perspectiva multivariante. Para llevar a cabo el trabajo se utilizan los análisis estadísticos Análisis Factorial y Análisis de Correspondencias.

2. MARCO TEÓRICO

Debido a la pandemia mundial producida por el COVID-19, España vivió alrededor de 3 meses de cuarentena, dependiendo de la provincia, provocados por un Estado de Alarma en los cuales hubo restricciones en los movimientos y salidas del domicilio. Este periodo provocó un cambio en la población, siendo más notado en el modo de trabajo, las compras y las relaciones sociales.

El día 14 de marzo de 2020 el Consejo de ministros aprobó la declaración del estado de alarma en todo el territorio español durante 15 días naturales con medidas de restricción al movimiento de personas, estando permitida la salida del domicilio personal únicamente para actividades esenciales (trabajo, compra, cuidado de personas dependientes...). El día 13 de abril volvieron al trabajo los empleados de la industria y la construcción y el 4 de mayo comenzaron las cuatro fases de la desescalada independientes en cada comunidad en función de la proporción de individuos contagiados.

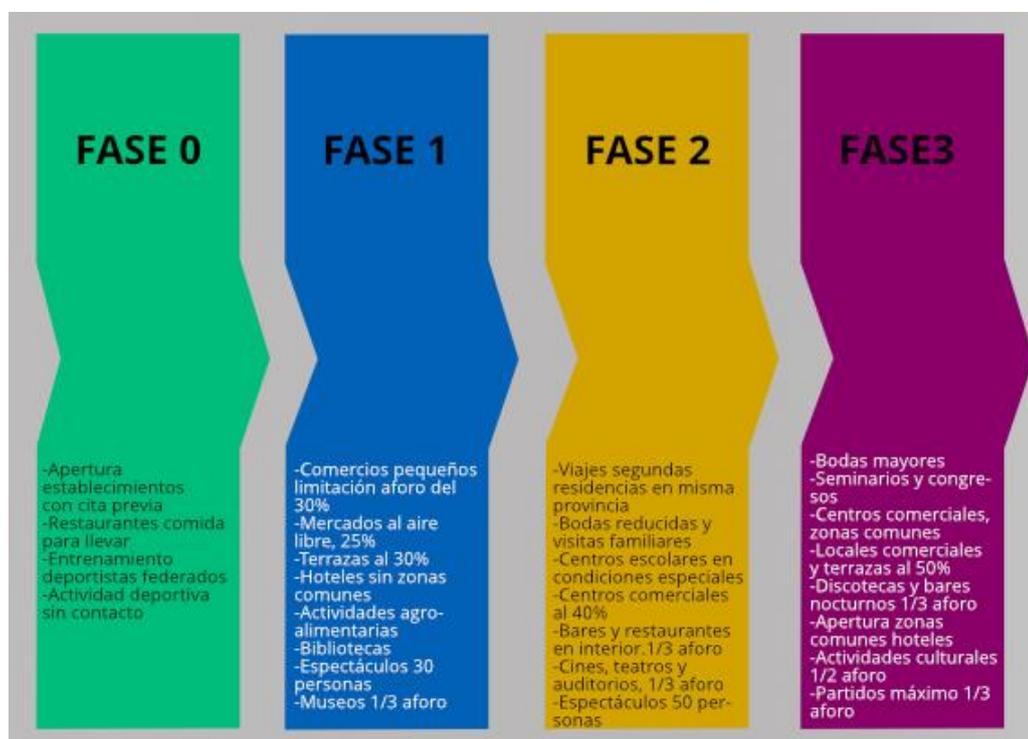


Ilustración 1. Fases de la desescalada presentadas por Pedro Sánchez.

Fuente: <https://psoegijon.es/las-cuatro-fases-del-plan-de-desescalada-presentado-por-pedro-sanchez/>

La aparición del virus provocó una crisis muy grande debido al estado de alarma que obligó a cerrar negocios, restringir la movilidad, realizar los estudios de manera online y la adaptación al teletrabajo. Esta crisis dejó a muchas familias sin dinero para poder llegar a fin de mes.

Durante el estado de alarma el gobierno obligó a realizar las compras en los establecimientos más cercanos a la residencia para así evitar los desplazamientos innecesario. La mayoría de los establecimientos de compra modificaron sus costumbres, cambiaron el

horario, daban preferencia a personas mayores de 65 años, pedían el uso de mascarilla, de guantes y de gel hidroalcohólico, una distancia de seguridad de 2 metros entre personas y que fuera solo una persona de cada unidad familiar a realizar la compra, evitando que fuera una persona mayor o de riesgo.

Jorge B., publica un artículo en la revista digital *Food Retail* en el que señala que durante la pandemia se pueden distinguir tres etapas:

1. La etapa de la “Preparación” corresponde a los días en los que ocurrieron los primeros casos positivos en España. En este tiempo el número de visitas al supermercado y la cantidad de compra no se vio modificada con respecto a los días anteriores.
2. La etapa de “Abastecimiento” corresponde con los días posteriores al aviso de un posible estado de alarma. Aumentó el número de visitas por día al supermercado y los carros volvían llenos de provisiones a las casas.
3. La última etapa, la de “Confinamiento” produjo otro cambio en el comportamiento del consumidor. Las compras se realizaban con menos frecuencia, para evitar las visitas al supermercado y por esto se seguían haciendo compras de gran tamaño. Eran compras planificadas, incluso planificando las zonas que visitar en el supermercado para pasar el menor tiempo posible.



Ilustración 2. Dinero gastado por minuto en cada etapa

Fuente: Bueno, J. (2020). *Análisis del comportamiento del shopper durante la Covid-19* [Ilustración]. Recuperado de https://www.foodretail.es/blogs/firma_invitada/Analisis-comportamiento-shopper-Covid-19_7_1460323955.html

En la *Ilustración 2*, obtenida del artículo *Análisis del comportamiento del shopper durante la Covid-19* se observa el cambio, en las tres etapas, del gasto por minuto. Las compras se caracterizaban por tener más productos en el carro en el mismo tiempo de visita.

Con motivo del miedo producido por el virus, las compras se comenzaron a hacer con el objetivo de llenar las casas, sin tener en cuenta las cantidades, que muchas veces eran desorbitadas. Como dijo Patricia Daimiel en su artículo “los seis comportamientos del consumidor ante el Covid-19”, se trataba de una cesta de “búnker” por lo que pueda pasar y durante el tiempo que haya que aguantar. A medida que iba avanzando el virus la población abastecía sus despensas de productos no perecederos como conservas, pasta, arroz, leche, harina y papel higiénico, además de productos de limpieza y desinfección.

Debido al gran número de compras online para evitar las grandes colas que había en ese momento y el riesgo al contagio, muchos supermercados y tiendas locales también modificaron

sus métodos de entrega. Había un plazo muy grande porque eran muchas las compras que había que entregar y algunos daban prioridad a personas mayores o de riesgo que vivían solas.

En un artículo publicado del periódico El País, Javier A. señala que los consumidores han abrazado el comercio online, no son fieles a una marca, son más impacientes y realizan menos compras y más grandes: “los consumidores ahora van con menos frecuencia a hacer la compra por seguridad y llenan más el carro, en el que priman los productos frescos de proximidad” (Fernández, 2021).

Uno de los cambios más grandes que ha producido la pandemia es la utilización masiva de los aparatos electrónicos. “El cambio ha involucrado a todos los segmentos de edad. Hemos enseñado a nuestros mayores a hablar por Zoom con sus nietos, pero también a hacer la compra por internet”, explica Patricia Daimiel, directora general de Nielsen. Las compras pasas a ser online en un porcentaje muy alto, todo por evitar colas, por tener distanciamiento social e incluso por reducir gastos.

El cambio en los consumidores no se produjo solo en cómo compraban si no en qué compraban. Según Ericson D. S., en su artículo *¿Cuál es el comportamiento del consumidor después del COVID-19?*, los consumidores comenzaron a adquirir más alimentos orgánicos por la realización de más comidas en casa y menos desperdicio de comida. Debido a las recomendaciones de la OMS (Organización Mundial de la Salud), numerosos consumidores adquirían cantidades de productos de higiene y salud, para reforzar la defensa ante el virus. El teletrabajo se convirtió en la forma de trabajo de gran cantidad de españoles, lo cual generó más ventas en tecnología.

Hablando del futuro, debido a la gravedad de la crisis vivida, se puede decir que los cambios que han sufrido los consumidores en su comportamiento de compra se mantendrán durante varios meses. Es imposible pensar que la pandemia y la situación que ha producido no dejará marcas en muchas familias. “De hecho, con el paso de la pandemia, queda claro que las necesidades han cambiado casi en su totalidad, siendo la salud y el bienestar, unos de los aspectos a tener en cuenta a la hora de ofrecer un servicio.” (Da Silva, 2020).

Es de esperar que tras finalizar la pandemia se mantengan formas de vida que se han adoptado durante su proceso, como la inclinación por el comercio online, la preocupación por la salud y la influencia de los valores en las decisiones de compra. Todo esto teniendo en cuenta la crisis económica que ha reducido los presupuestos de numerosas familias y que ha incrementado la importancia del precio (Flavián, 2021).

3. OBJETIVOS

Conocer como es el comportamiento de los consumidores de los supermercados debido a la pandemia producida por la Covid-19.

Medir la fiabilidad cuestionario y analizar la estructura factorial del cuestionario utilizado para medir el comportamiento de los consumidores en los supermercados.

Caracterizar a los consumidores y conocer cómo se comportan en función de si sus hábitos se han visto modificados o no por la actual pandemia.

4. METODOLOGÍA

4.1 DESCRIPCIÓN DE LOS DATOS

Los datos para el estudio han sido obtenidos a partir de un cuestionario sobre el consumo de los individuos en los supermercados durante la Covid'19, el cual se puede observar en el *Anexo I*. Dicho cuestionario consta de 29 preguntas categóricas de las cuales ocho son preguntas sociodemográficas y el resto están divididas en cuatro partes.

La primera parte es la correspondiente con el “supermercado” y agrupa las preguntas relacionadas con el supermercado habitual y el motivo de haberlo cambiado, si lo hay. La segunda parte corresponde con las preguntas de “compra”, es decir, el número de compras al mes, tipo de alimentos en las compras, planificación y días de la compra, etc. La tercera parte se centra en la “economía” del individuo, queriendo saber, si se invierte el mismo dinero en las compras, si se espera que vuelvan a cambiar los ingresos, etc. Y por último la parte correspondiente a las preguntas relacionadas con “compra online”, para saber si se han introducido las compras online tras empezar la pandemia, en qué cantidad, los problemas para realizarlas, etc.

La muestra se tomó entre los días 15 de febrero y 9 de marzo de 2021 y la población objetivo de estudio ha sido la población mayor de 18 años de la región española que realiza la compra en al menos un supermercado. Debido a la situación vivida desde hace más de un año la obtención de datos ha sido muy complicada. La encuesta ha sido realizada de manera *Online* a través de un formulario de *Google Forms* para evitar el contacto físico y poder acceder a más individuos, lo que sesga mucho la muestra porque no se consigue el mismo número de datos de todos los intervalos de edad. A pesar de esto, se obtuvo información sobre 371 individuos pertenecientes a 39 provincias.

El método de muestreo utilizado es un muestro mixto en el que se utiliza el muestreo intencional o de conveniencia y el muestreo por bola de nieve. En un principio se utiliza el muestreo intencional porque se buscan individuos de los cuales ya se conoce que realizan la compra en el supermercado y a partir de ahí, estos individuos, son los que van encontrando otros con la misma condición.

Para llevar a cabo los respectivos análisis se han utilizado los programas IBM SPSS Statistics 22 e IBM SPSS Amos 26.

4.2 ANÁLISIS ESTADÍSTICOS

Para realizar el trabajo se han utilizado tres técnicas estadísticas. En primer lugar, se analiza la fiabilidad del cuestionario, para medir la confianza concedida a los datos obtenidos con el mismo.

En segundo lugar, se realiza un Análisis Factorial para conocer la estructura interna del cuestionario y si es posible, reducir su dimensión.

Y, por último, se lleva a cabo un Análisis de Correspondencias Simples, para caracterizar a los consumidores y crear distintos perfiles.

4.2.1 FIABILIDAD

La fiabilidad de un test hace referencia a la precisión que tiene el test, utilizado como instrumento de medida. Además, se considera que una medida es fiable si está relativamente libre de errores de medida. Se puede entender la fiabilidad de tres formas distintas:

- Fiabilidad intra. Se refiere a la consistencia interna, al nivel en que los ítems están relacionados entre sí y se mide mediante el Alfa de Cronbach.

$$\alpha_{Cronbach} = \frac{m}{m-1} \left(1 - \frac{\sum var(x_i)}{var(S)} \right) \quad (2)$$

Donde m es el número de ítems de una dimensión, $var(x_i)$ es la varianza del i -ésimo ítem de la escala y $S = \sum x_i$. Si los ítems no están correlacionados, el valor de Alfa será cero y, si los ítems son idénticos y tienen correlación perfecta, el valor de Alfa será uno. Se acepta el valor de Alfa cuando está por encima de 0,7.

- Fiabilidad test-retest. Hace referencia a la estabilidad de los valores de la escala al aplicarla en tiempos distintos. Se mide utilizando el Coeficiente de Correlación de Pearson o el Coeficiente de Correlación Intraclase, definido como

$$CCIntra = \frac{var(i)}{var(t)} = \frac{var(i)}{var(i) + var(e)} \quad (3)$$

Donde $var(i)$ es la varianza entre los individuos y $var(t)$ es la varianza total, la cual incluye a $var(i)$ y $var(e)$, que es la varianza debida a los errores aleatorios de medida. Para variables continuas este coeficiente es equivalente al coeficiente *Kappa*.

- Fiabilidad inter. Describe la consistencia entre la respuesta y de distintos individuos con respecto a la misma escala. Algunos autores definen esta fiabilidad como

$$Fiabilidad = \frac{var(diferencias\ entre\ sujetos)}{var(t)} \quad (4)$$

Los coeficientes de fiabilidad tienden a ascender cuando la muestra es heterogénea y grande y cuando hay muchos ítems con una formulación muy semejante y con más de dos respuestas.

4.2.2 ANÁLISIS FACTORIAL (AF)

El Análisis Factorial es una técnica estadística multivariante que representa las relaciones entre un conjunto de variables y permite reducir el tamaño de los datos sin perder demasiada información.

El origen de esta técnica se encuentra en las técnicas de regresión lineal realizadas por Galton. Más tarde fue K. Pearson quién en 1901 presentó la primera propuesta del método de componentes principales. Se atribuye a Charles Spearman en 1904 el origen del AF en un trabajo sobre la inteligencia en el cual distinguió un factor general y otros factores específicos. A las ideas de Spearman se añadieron más tarde aportaciones de otros investigadores.

Se supone el estudio de un problema con X_1, X_2, \dots, X_n variables en el que, si n es un valor muy alto, el estudio será muy complicado y además no todas las variables aportarán información relevante e incluso pueden medir conceptos similares. El AF sirve por tanto para trabajar con un número de variables más pequeño.

Antes de aplicar el AF hay que saber si los datos son válidos y para ello hay que tener en cuenta que las variables han de estar altamente correlacionadas y la matriz de correlación ha de tener una cierta estructura. Para comprobar esto, se utilizan las siguientes técnicas:

- La forma más primitiva es observando la matriz de correlaciones, pero, si el número de variables es grande, es difícil obtener conclusiones porque dicha matriz será muy grande y será muy difícil visualizarla en su conjunto.
- El determinante de la matriz de correlaciones ha de ser bajo. Si ocurre esto se puede aplicar el AF y, además, cuanto más cercano sea a 0 el determinante, mayor multicolinealidad.
- El test de esfericidad de Bartlett, en el cual se contrasta la hipótesis nula de que las variables analizadas no están correlacionadas en la muestra, es decir, que la matriz de correlaciones es la matriz identidad. Este test es muy bueno cuando el tamaño muestral es pequeño.
- La medida de Kaiser – Meyer – Olkin (KMO), la cual indica qué tan apropiados son los datos para realizar el análisis. Mide lo adecuado que es el muestreo para cada variable en el modelo y para el modelo completo. Esta medida compara las magnitudes de los coeficientes de correlación parcial, de forma que cuánto más pequeño sea su valor, mayor será el valor de los coeficientes de correlación parciales y como consecuencia, menos apropiado será realizar un AF. Se considera un valor KMO aceptable a partir de 0,5.
- Si el coeficiente de correlación múltiple es alto, hay una alta adecuación al AF.
- La matriz de correlación anti-imagen está formada por los negativos de los coeficientes de correlación parcial. Si los elementos que no están en la diagonal son pequeños, se puede aplicar el AF

El esquema de los pasos de un AF es el siguiente:

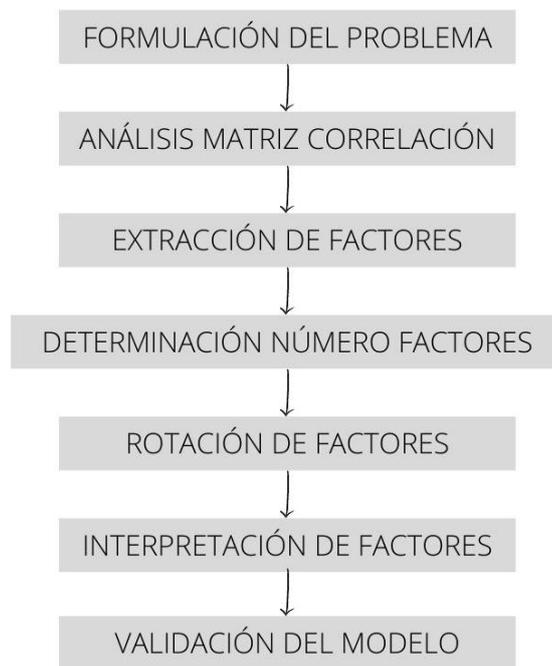


Ilustración 3. Esquema Análisis Factorial

El AF plantea que las relaciones entre las variables pueden ser explicadas mediante otras variables no observadas o latentes llamadas factores y comienza con una matriz de datos con las observaciones de j variables para n individuos.

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1j} \\ x_{21} & x_{22} & \dots & x_{2j} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nj} \end{pmatrix} \quad (5)$$

El modelo del AF viene dado por las siguientes ecuaciones:

$$\begin{aligned} X_1 &= \alpha_{11}F_1 + \alpha_{12}F_2 + \dots + \alpha_{1m}F_m + U_1 \\ X_2 &= \alpha_{21}F_1 + \alpha_{22}F_2 + \dots + \alpha_{2m}F_m + U_2 \\ &\dots \\ X_j &= \alpha_{j1}F_1 + \alpha_{j2}F_2 + \dots + \alpha_{jm}F_m + U_j \end{aligned} \quad (6)$$

El modelo tiene j variables observables (X_1, \dots, X_j), m variables no observables llamadas factores comunes (F_1, \dots, F_m) y j variables no observables llamadas factores únicos o específicos (U_1, \dots, U_j). Estos factores únicos explican la varianza de la variable j -sima que no es explicada por los factores comunes. Los coeficientes α_{ij} son las saturaciones de la variable X_i en el factor F_j , también se pueden llamar pesos, cargas o ponderaciones.

El modelo supone que las variables y los factores están estandarizados, es decir, que tienen media 0 y varianza 1. Supone también que los factores comunes y únicos están incorrelacionados. Además, m será menor que j ya que se pretende explicar las variables observadas con un menor número de otras hipotéticas.

Por tanto, el objetivo del AF es buscar y encontrar una matriz A tal que sus elementos sean α_{ij} y para ello existen varias técnicas que se explicarán más adelante.

Los factores son un rasgo hipotético que están subyacentes de una variable y que influyen en el comportamiento que tienen los ítems de un test. El vocablo “factor” hace referencia a la variable teórica, procedente de un AF a partir de la matriz de intercorrelaciones.

El AF asigna “pesos” a cada variable. Estos pesos factoriales indican la carga que tiene cada variable en cada factor y lo ideal es que cada variable cargue alto en un factor y bajo en los demás. El peso se infiere analizando lo que tienen en común las variables que están agrupadas en un mismo factor. Para interpretar el significado de un factor son necesarios dos pasos:

- Determinar los pesos más altos en valor absoluto. Hay que fijarse en los pesos, en valor absoluto, que son mayores que un valor crítico establecido. Para facilitar la interpretación, se ordenan los pesos y se destacan los que son más relevantes.
- Dar nombre a los factores. Es importante poner una etiqueta a los factores que refleje lo que las variables que más carga tienen en él tienen en común. Las variables que más importancia han de tener en la elección del nombre son las que más pesen.

La interpretación de la estructura factorial puede ser complicada para el investigador. Puede ocurrir por ejemplo que las variables que se han agrupado en un factor son muy distintas y no tienen un argumento común. También puede pasar que las variables tengan saturaciones altas en más de un factor y, por tanto, que midan más de una característica y sea más difícil entender su significado.

COMUNALIDAD Y UNICIDAD

Del modelo factorial nombrado anteriormente se obtiene la ecuación fundamental $X_j = \sum_{m=1}^M \alpha_{jm} F_m + U_j$, donde las m variables se describen linealmente en función de las m componentes incorrelacionadas F_1, F_2, \dots, F_n y a partir de esta ecuación y de las propiedades de las combinaciones lineales de variables, se puede expresar la varianza de las variables observadas como:

$$\sigma_{X_j}^2 = \sum_{m=1}^M \alpha_{jm}^2 \sigma_{F_m}^2 + \sum_{m=1}^M \sum_{\substack{m'=1 \\ m' \neq m}}^M \alpha_{jm} \alpha_{jm'} \rho_{F_m F_{m'}} + \sigma_{E_j}^2 \quad (7)$$

Donde $\rho_{F_m F_{m'}}$ es la covarianza entre los factores F_m y $F_{m'}$. Teniendo en cuenta que los factores tanto únicos como comunes están tipificados, sus varianzas serán unitarias y las covarianzas coincidirán con los coeficientes de correlación y de esta forma $\sigma_{X_j}^2 = 1$. En la *Tabla 1* se observa la descomposición de la varianza de una variable observada.

Varianza de la variable observada	Varianza debida a los factores comunes	Unicidad / Varianza específica / Error
$\sigma_{X_j}^2 = 1$	$\sum_{m=1}^M \alpha_{jm}^2 + \sum_{m=1}^M \sum_{\substack{m'=1 \\ m' \neq 1}}^M \alpha_{jm} \alpha_{jm'} \rho_{F_m F_{m'}}$	$\sigma_{E_j}^2 = u_j^2 \sigma_{U_j}^2 + u_j \sum_{p=1}^m \alpha_{jp} \sigma_{F_p} U_j$

Tabla 1. Descomposición de la varianza de una variable observada

De la descomposición de la varianza de cada variable se obtienen dos conceptos fundamentales en el AF:

- Comunalidad. Es la proporción de la varianza de las variables originales que es explicada por los factores comunes y se representa por h_j^2 . Dicho de otra forma, es la suma de los pesos factoriales al cuadrado en cada una de las filas del modelo factorial y se obtiene de la siguiente expresión:

$$h_j^2 = \frac{\sum_{m=1}^M \alpha_{jm}^2 + \sum_{\substack{m'=1 \\ m' \neq 1}}^M \alpha_{jm} \alpha_{jm'} \rho_{F_m F_{m'}}}{\sigma_{X_j}^2} = \sum_{m=1}^M \alpha_{jm}^2 + \sum_{\substack{m'=1 \\ m' \neq 1}}^M \alpha_{jm} \alpha_{jm'} \rho_{F_m F_{m'}} \quad (8)$$

La proporción de varianza que explican los factores depende del peso factorial de la variable en cada factor y de las correlaciones entre los factores correspondientes. Por hipótesis los factores únicos suelen estar incorrelacionados con los comunes, de esta forma $\rho_{F_m F_{m'}} = 0$ para todo m y m' y, por tanto, el último sumando de la expresión anterior se anula. Por otro lado, si los factores comunes están incorrelacionados entre sí, la expresión anterior se reduce a:

$$h_j^2 = \sum_{m=1}^M \alpha_{jm}^2 \quad (9)$$

- Unicidad. Corresponde a la varianza específica o de error. Se simboliza mediante u_j^2 e indica la proporción de varianza de la variable j que no depende de factores comunes, es decir, que viene expresada por el factor único. Se obtiene a partir de:

$$u_j^2 = \frac{\sigma_{Ej}^2}{\sigma_{Xj}^2} = 1 - h_j^2 \quad (10)$$

Volviendo a la descomposición de la *Tabla 1* se puede escribir

$$\sigma_{Xj}^2 = h_j^2 + u_j^2 = 1 \quad (11)$$

Igual que se puede obtener la comunalidad (proporción de varianza de una variable explicada por los factores comunes) también se puede obtener la proporción de varianza de todas las variables explicada por los factores comunes, que corresponde a la varianza total (VT) y se define como la suma de las varianzas de todas las variables.

$$VT = \sum_{j=1}^J \sigma_{Xj}^2 = \sum_{j=1}^J (h_j^2 + u_j^2) = \sum_{j=1}^J h_j^2 + \sum_{j=1}^J u_j^2 \quad (12)$$

Según Galton, en sus trabajos de 1889, que dos variables estén relacionadas entre sí se debe al hecho de que ambas poseen algo en común y algo que las diferencia. Así, la varianza total de una variable se debe a factores que comparte con otras variables (comunalidad) y a factores específicos y únicos de la variable (unicidad).

Si las puntuaciones en los ítems están tipificadas, la parte de la varianza total que corresponde a los factores comunes es

$$VC = \sum_{j=1}^J h_j^2 \quad (13)$$

y la proporción de varianza total explicada entre todos los factores comunes es

$$PVC = \frac{VC}{VT} \quad (14)$$

Cuanto más se acerque el valor de PVC a 1, más poder explicativo tendrán los factores comunes. Si los factores son independientes, la fórmula de la comunalidad se simplifica y su proporción se puede descomponer para obtener la proporción de varianza explicada de cada factor por separado. Es decir, la proporción de varianza total explicada por el factor m puede ser calculada como

$$PVC_{Fm} = \sum_{j=1}^J \frac{\alpha_{jm}^2}{J} \quad (15)$$

En la *Tabla 2* aparece la proporción de varianza explicada por cada factor común independiente.

Prop. Varianza total explicada por factores comunes	Prop. varianza total explicada por el primer factor	...	Prop. Varianza total explicada por el factor m	...	Prop. Varianza total explicada por el factor M
PVC	PVC_{F1}	...	PVC_{Fm}	...	PVC_{FM}
$\frac{\sum_{j=1}^J h_j^2}{J} =$	$\frac{\sum_{j=1}^J \alpha_{j1}^2}{J} +$...	$\frac{\sum_{j=1}^J \alpha_{jm}^2}{J} +$...	$\frac{\sum_{j=1}^J \alpha_{jM}^2}{J}$

Tabla 2. Proporción de varianza explicada por cada factor común independiente

El valor de la proporción de varianza total explicada por el factor m (PVC_{Fm}) determina la importancia que tiene el factor ya que los diferentes factores que se extraen no tienen todos la misma importancia porque cada uno explica una cantidad concreta, expresada porcentualmente, de la varianza total de las variables. Que un factor explique un porcentaje alto de la varianza total, es un indicio de que las cargas de las variables en ese factor son altas e incluso puede ser que haya muchas variables que tienen peso en ese factor, lo cual significa que, a la hora de describir las relaciones entre las variables originales, es una dimensión importante.

4.2.2.1 ANÁLISIS FACTORIAL EXPLORATORIO (AFE)

El Análisis Factorial Exploratorio (AFE) es un conjunto de métodos estadísticos multivariados de interdependencia (todas las variables son analizadas en conjunto, no hay variable independiente y variables respuesta) que tienen por objetivo identificar una estructura de factores subyacentes a un conjunto de datos. Es una técnica muy utilizada en las ciencias sociales pero su mayor notabilidad es en el campo de la psicometría. Este método permite agrupar las variables que están fuertemente correlacionadas entre sí.

Anteriormente se nombró al estadístico Galton y fue él quien en 1889 afirmó que había que desarrollar una técnica para descubrir los factores o variables latentes subyacentes. El comienzo del AFE también puede atribuirse a Pearson (1901) porque desarrolló el coeficiente de correlación y planificó los principios en los que se basa el AFE. Sin embargo, se puede considerar también al psicólogo Spearman (1904) como el creador de este análisis cuando lo empleó para estudiar las correlaciones entre diferentes pruebas de habilidades. Con el paso de los años han sido numerosos autores los que han continuado con la tarea de desarrollar la técnica del AF.

Los pasos a seguir para realizar el AFE son los siguientes:

- i. El investigador debe decidir el número m que corresponderá con el número de factores que se van a extraer y el que será el método de extracción, que es el método de estimación de los pesos.

- ii. El siguiente paso es analizar la solución de un factor. Si el ajuste de los datos es bueno, se puede concluir que la solución es unidimensional y, si no, se irá probando sucesivamente hasta tener los factores suficientes para el modelo.
- iii. A la solución obtenida se la denomina matriz factorial inicial, la cual es difícil de interpretar y además no es la única solución posible.
- iv. Se pasa al método de rotación, en donde el investigador decide si realizar una rotación ortogonal o una rotación oblicua, con el objetivo de facilitar la interpretación.
- v. Por último, se interpretan los resultados, indicando las variables que corresponden a cada factor y el nombre para cada uno de ellos.

EXTRACCIÓN DE FACTORES

Antes de comenzar con el AFE habría que comprobar si es una buena solución para analizar los datos, con los métodos que se explicaron en el AF.

El método de extracción de factores es el procedimiento para estimar los parámetros del modelo. Puesto que los cálculos son complicados se realizan con programas informáticos como SPSS o R. Muchos de los algoritmos que se emplean para encontrar la mejor solución, son métodos iterativos, es decir, parten de una solución inicial y en los siguientes pasos van mejorando la solución hasta llegar a la óptima, en caso de que sea posible. Los métodos utilizados con más frecuencia son los siguientes:

- Método de Componentes Principales. Consiste en estimar las puntuaciones factoriales mediante puntuaciones tipificadas de las primeras k-componentes y la matriz de pesos, mediante las correlaciones de las variables originales con dichas componentes. Este método explica la mayor cantidad posible de varianza en los datos observados y, por tanto, analiza la varianza total asociada a las variables, incluyendo la varianza específica y la varianza de error. El primer componente tiene la varianza máxima y las sucesivas explican progresivamente proporciones menores de la varianza.

La ventaja de este método es que siempre proporciona una solución y su inconveniente es que al no estar basado en el modelo de AF puede proporcionar estimadores muy sesgados de la matriz de pesos y además no se puede utilizar si la matriz de correlaciones es singular.

- Método de Factorización de Ejes Principales. Solo contempla la varianza que las variables tienen en común, es decir, la covarianza, excluyendo la varianza específica y la atribuible al error de medida. Parte de la matriz de correlaciones original con los cuadrados de los coeficientes de correlación múltiple puestos en la diagonal principal como estimaciones iniciales de las comunalidades. Las cargas que resultan se utilizan de nuevo para estimar las comunalidades que

reemplazan a las estimaciones previas en la diagonal. El proceso continúa hasta satisfacer el criterio de convergencia para la extracción.

Su ventaja es que recupera factores débiles y es buena opción si hay pocas variables. La desventaja de su uso es que no proporciona índices de ajuste ni intervalos de confianza.

- Método de Mínimos Cuadrados no Ponderados. Para un número fijo de factores genera una matriz de coeficientes que minimiza la suma de las diferencias al cuadrado entre las matrices de correlación observada y reproducida, eliminando en las diferencias los elementos de la diagonal.
- Método de Mínimos Cuadrados Generalizados. Minimiza lo mismo que el método anterior, pero ponderando las correlaciones inversamente por la varianza del factor específico. Permite aplicar contraste de hipótesis para determinar el número de factores.
- Método de Máxima Verosimilitud. Proporciona las estimaciones de los parámetros que han producido la matriz de correlaciones observadas, asumiendo que la muestra proviene de una distribución normal multivariante y además no depende de la escala de medida de las variables.
Su principal beneficio es que permite estimar la significación estadística de las cargas factoriales y que genera intervalos de confianza con dichas estimaciones. Su inconveniente es que al usarse solo cuando los datos siguen una distribución normal multivariante hace que en la práctica sea uno de los métodos menos utilizados.
- Método de Factorización por Imágenes. Este método fue desarrollado por Guttman y se basa en la teoría de las imágenes. La imagen parcial es la parte común de una variable y esta se define como su regresión lineal sobre las variables restantes.
- Método Alfa. Considera las variables que se incluyen en el modelo como una muestra del universo de las posibles variables. Maximiza el Alfa de Cronbach para los factores.

¿Cómo saber que método utilizar? No hay una respuesta concreta sobre qué modelo utilizar porque depende de distintos elementos como el tamaño de la muestra, la distribución de los datos, el objetivo del estudio, etc.:

- Si las comunalidades son mayores que 0,6, todos los métodos tienden a dar la misma solución.
- Si las comunalidades son bajas en alguna de las variables, el método de componentes principales tiende a dar soluciones con cargas factoriales mayores que el resto de los métodos.

- Si la base de datos consta de más de 30 variables todos los métodos tienden a ofrecer el mismo resultado porque las estimaciones de la comunalidad tienen menos influencia en la solución obtenida. Aun así, el método más fácil cuando se analiza un conjunto con numerosas variables es el método de componentes principales.
- El método más popular y que incluyen todos los paquetes de programación por defecto es el método de componentes principales. Es el más útil si el objetivo es reducir la información o establecer una teoría sobre los datos.
- Si se sabe que los datos son extraídos de una población normal multivariante y que no tienen casos atípicos la mejor opción es el método de máxima verosimilitud. Es útil cuando se necesitan índices sobre la bondad de ajuste del modelo.
- Si los datos no siguen la distribución normal multivariante y el objetivo es obtener una solución teórica que no esté contaminada por la varianza específica y la de error, el método que más conviene es el método de ejes principales.

NÚMERO DE FACTORES

Una vez decidido el método de extracción de factores hay que decidir el número de factores que van a ser seleccionados. Es una decisión importante puesto que el objetivo es la reducción de la dimensión sin perder demasiada información.

El autovalor o valor propio de un factor es la magnitud de varianza explicada por ese factor. Se obtiene de la varianza explicada de una variable, haciendo el cuadrado de la correlación entre la variable y un factor y, sumando todos los cuadrados de los pesos factoriales de las variables en un factor, se obtienen los valores propios.

Hay varios criterios para elegir el número de factores a utilizar:

- Regla de Kaiser. Sugiere que se deben usar los factores cuyo valor propio es igual a uno. El razonamiento para que el punto de corte sea uno es que las variables están estandarizadas con varianza igual a 1 y no es adecuado interpretar un factor que explique menos varianza que la explicada por una variable particular.

Dividiendo el autovalor por el número de variables y multiplicándolo por 100 se obtiene el porcentaje de varianza que explica ese factor.

El inconveniente es que suele conducir a extraer demasiados factores, sobre todo si los datos constan de 50 o más variables.

- Regla del 75% de la varianza. Se determina el número de factores según la absorción de inercia, es decir, serán suficientes factores cuando el porcentaje de la varianza total absorbida sea de 75%.

- Regla del codo. Se realiza analizando el *Scree Plot*, un gráfico de la magnitud de los autovalores en el que se representan los autovalores en el eje de ordenada y el número de factores en el de abscisa. El nombre de *scree* se debe al parecido que tiene con el perfil a la falda de una montaña. Esta técnica selecciona un grupo de factores que tengan valores propios significativamente superiores a los demás. Para determinar esto, en el gráfico, se identifica un punto de inflexión en la curva del gráfico que suele estar determinado por el primer cambio de pendiente en la gráfica y a partir de la cual la curva se va transformando en una línea relativamente recta. Los autovalores residuales se encuentran a la derecha del gráfico, formando casi una línea recta y los que explican la mayor parte de la varianza se encuentran en la parte izquierda, formando una gran pendiente. El nombre de la regla se debe a la semejanza que forman los puntos y las líneas con un brazo.
- Regla del análisis paralelo. Esta regla suele complementar a las anteriores cuando el número de variables iniciales y de factores es elevado. En este análisis se generan autovalores de una matriz de datos aleatorios pero que contiene el mismo número de variables y casos que la matriz original. Considera que los factores a extraer deben dar cuenta de más varianza que la que se espera de manera aleatoria. Los factores que tienen valores propios mayores a los valores aleatorios se retienen para la interpretación.
- Especificación por el investigador. Si existe una teoría sobre el número de factores, el investigador determinará a priori ese número como factores a retener. Es el criterio más fiable si los datos y las variables están bien elegidos.
- Criterio de división a la mitad. Se divide la muestra en dos submuestras iguales al azar y se realiza el AF en ambas. Se conservan los factores que tienen alta correspondencia de cargas factoriales en las dos muestras. Antes de realizarlo se debe comprobar que no existen diferencias significativas en cuando a las variables estudiadas se refiere.

Decidir el número de factores que son extraídos es una decisión problemática dentro del AF. Emplear solamente un criterio de los anteriores puede llevar a sobrestimar o subestimar el número de factores reales, por ello, se recomienda emplear más de un criterio para identificar el número de factores subyacentes. Eligiendo entre incluir más o menos factores, la sobre – extracción tiene menos riesgos porque conduce a menos error en la medición.

ROTACIÓN DE FACTORES

Lo normal una vez obtenida una solución es que la matriz no defina unos factores que se puedan interpretar. El objetivo es conseguir que las saturaciones sean altas a costa de otras que serán bajas para poder destacar la influencia de los factores comunes sobre las variables observables.

Los métodos de rotación pretenden aproximar la solución factorial al *Principio de Estructura Simple*, según el cual la matriz de cargas factoriales debe cumplir tres características:

- i. Cada factor debe tener unos pocos pesos altos y los demás próximos a cero.
- ii. Cada variable no debe estar saturada más que en un factor.
- iii. No deben existir factores con la misma distribución. Es decir, dos factores distintos deben presentar distribuciones diferentes de cargas altas y bajas.

La rotación factorial consiste en hacer girar los ejes de coordenadas, que representan a los factores, hasta conseguir que se aproxime al máximo a las variables en las cuales están saturados.

Hay dos tipos de rotaciones en función de si los factores rotados seguirán siendo ortogonales o no y en ambos casos la comunalidad de cada variable no se modifica, es decir, la rotación no afecta a la bondad de ajuste de la solución factorial, pero sí cambia la varianza explicada por cada factor.

- Rotación Ortogonal. Se caracteriza por encontrar la estructura más simple bajo la restricción de que los factores sean ortogonales, es decir, no estén correlacionados. Los ejes se rotan de manera que quede resguardada la incorrelación entre los factores. Los nuevos ejes son perpendiculares a los ejes sin rotar. Es adecuada cuando el método de extracción de factores es el método de componentes principales. Existen tres tipos de rotación ortogonal:
 - Rotación *Varimax*. Minimiza la variabilidad de los coeficientes para cada factor y así reduce el número de variables que tienen saturaciones altas en cada factores. Simplifica la interpretación de los factores mejorando la solución por columna. Es el método más empleado si el número de factores es pequeño.
 - Rotación *Quartimax*. Minimiza la variabilidad de los coeficientes para cada variable y así reduce el número de factores para explicar cada variable. Simplifica la interpretación de las variables mejorando la solución por filas. Es el método más empleado cuando el número de factores es alto.
 - Rotación *Equamax*. Es una mezcla de la rotación *Quartimax*, que simplifica las variables, y la *Varimax*, que simplifica los factores. Con este método se minimiza el número de variables que saturan alto en un factor y el número de factores que hacen falta para explicar una variable.
- Rotación Oblicua. Se aplica cuando se supone que los factores en la población están fuertemente correlacionados, la matriz de rotación no tiene la exigencia de ser ortogonal, basta con que sea no singular.
 - Rotación *Oblimin directo*. Es el método de rotación oblicua más utilizado e intenta encontrar una estructura simple teniendo en cuenta que las saturaciones ya no representan la correlación entre los factores y las variables.

Necesita un valor δ , preferentemente entre -0,5 y 0,5, que controla el grado de correlación entre los factores.

- Rotación *Promax*. Altera los resultados de una rotación ortogonal hasta crear una solución con saturaciones lo más próximas a la estructura ideal. Esta estructura se supone que se obtiene al elevar las saturaciones que se obtienen de una rotación ortogonal a una potencia normalmente entre 2 y 4. Cuanto mayor sea la potencia, más oblicua es la solución.

La ortogonalidad absoluta es únicamente teórica y, a fines prácticos, se entiende que una solución es ortogonal cuando todas las correlaciones entre los factores son inferiores a 0,32. En 2001 Tabachnick y Fidell propusieron la realización de una rotación oblicua inicial como filtro para obtener la matriz de correlación entre factores. Si se observa alguna correlación entre los factores superiores a 0,32, se debe escoger una rotación oblicua y, en caso contrario, una ortogonal.

INTERPRETACIÓN DE LOS RESULTADOS

La última tarea del AFE es interpretar los factores. Para realizar este paso, se observa el patrón de correlaciones bajas y altas de cada variable en los distintos factores en la matriz rotada.

Se recomienda que cada factor tenga, al menos, cuatro variables con correlaciones iguales o superiores a 0,40, para poder ser interpretado y que hay que tener en cuenta las correlaciones más altas para indicar el nombre de cada factor.

4.2.2.2 ANÁLISIS FACTORIAL CONFIRMATORIO (AFC)

El razonamiento del AFE y AFC son el mismo, pero con importantes diferencias. En el AFE se estudia la estructura que mejor se ajusta a los datos y es una técnica exploratorio porque no se estipula directamente un modelo teórico. De este modo, si el modelo de AFE que se obtiene no coincide con el modelo teórico, no significa que sea incorrecto. Para decidir esto, hay que aplicar el Análisis Factorial Confirmatorio (AFC).

Los pasos para realizar un AFC son los siguientes:

- i. Representación gráfica del modelo.
- ii. Estudiar si el modelo está identificado, es decir, si existe suficiente información para estimar sus parámetros. Si esto no ocurre, no se puede estimar y hay que redefinir el modelo.
- iii. Estimar los parámetros del modelo. Para este paso hay que escoger un método de estimación mediante el cual se obtienen los parámetros que minimizan las discrepancias entre las varianzas y covarianzas reproducidas por el modelo (S^*) y las observadas (S).

- iv. Obtener los indicadores de ajuste del modelo. Para esto se contrasta si los datos empíricos se ajustan al modelo teórico, es decir, si el parecido entre S y S^* es suficiente para confirmar que el modelo se ajusta a los datos.

Los modelos del AFC se simbolizan por un *path diagram* (diagrama de senderos), en la *Ilustración 3* obtenida del documento *Análisis factorial confirmatorio. Su utilidad en la validación de cuestionarios relacionados con la salud* se muestra a diferencia entre un diagrama de senderos para el AFE y otro para el AFC, donde se representa el modelo para las respuestas a 7 variables. Se observa que los ítems 1, 2 y 3 pertenecen a un factor y los ítems 4, 5, 6 y 7 a otro factor.

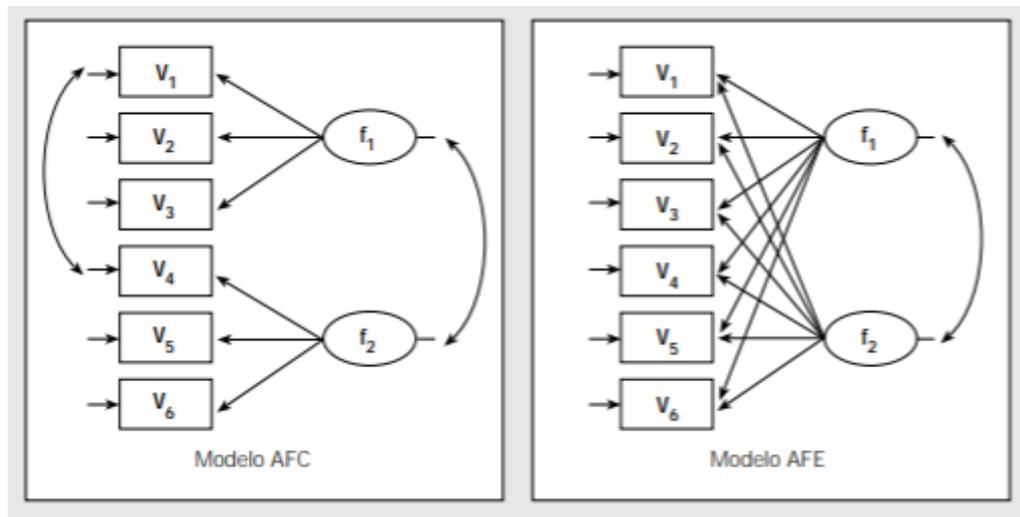


Ilustración 4. AFE y AFC

Fuente: Batista-Foguet, J., M., Coenders, G., Alonso, J. (2004). *Análisis factorial confirmatorio. Su utilidad en la validación de cuestionarios relacionados con la salud.* [Ilustración].

Las variables observables ($X_1, X_2, X_3, X_4, X_5, X_6, X_7$) se representan mediante cuadrados, las variables latentes como los factores (F_1 y F_2) o los errores (E_1, E_2, \dots, E_7) se representan mediante círculos. Las relaciones entre las variables se representan con líneas. Una línea curva bidireccional conectando dos variables latentes indica que ambas covarían. Las flechas rectas indican que hay una relación direccional entre las dos variables conectadas. Por ejemplo, X_1 recibe líneas de F_1 y de E_1 , lo que significa que está influenciada por ambas variables.

Las variables endógenas son las variables a las que les llega alguna flecha y las variables exógenas son a las que no les llega ninguna flecha. En los modelos de AFC, las variables latentes siempre son exógenas y las variables observables siempre son endógenas. Se asume que las variables exógenas y los errores no están correlacionados y esta condición se llama pseudo-aislamiento.

Los modelos pueden contener tres tipos de parámetros, según las especificaciones del investigador. Un parámetro libre es estimado sin condición alguna a partir del conjunto de datos. Un parámetro libre es definido para ser igual a una constante. Y, un parámetro

restringido, es estimado bajo alguna restricción, que suele estar relacionada con valores de otros parámetros restringidos.

IDENTIFICACIÓN DEL MODELO

Se considera un modelo no identificado cuando no es posible estimar sus parámetros, es decir, existen distintos conjuntos de parámetros que llevan al mismo ajuste estadístico. En el AFC el modelo debe estar identificado para que la solución factorial que se obtiene sea única. Para que el modelo sí esté identificado hay que tener en cuenta dos aspectos:

- Fijar la métrica de las variables latentes. Hay que fijar la escala de medida de los factores latentes, es decir, la desviación típica de esas variables. Para determinar la métrica existen distintas formas:
 - Fijar la varianza de los factores a uno, asumiendo entonces que los factores están estandarizados. La ventaja de hacerlo así es que las covarianzas entre factores son las correlaciones entre ellos.
 - Para cada factor, fijar el peso de una variable a uno, así la varianza del factor latente es función de la varianza de la variable medida.

Cuando se fija la métrica del factor se reduce el número de incógnitas del modelo y de esta forma, la cantidad de parámetros a estimar.

- Comprobar el número de grados de libertad, que ha de ser positivo. Hay que distinguir entre el número de datos disponibles $(J(J + 1))/2$ elementos disponibles de la matriz de varianzas – covarianzas, donde J es el número de variables) y el número de parámetros a estimar (t).

Para que un modelo esté identificado, el número de parámetros debe ser menos que el número de datos disponibles y, a la diferencia entre el número de datos y el de parámetros estimados se le denomina grados de libertad (gl)

$$gl = \frac{J(J + 1)}{2} - t \tag{16}$$

Los grados de libertad pueden ser:

- Negativos. Hay más parámetros a estimar que datos, el modelo no está identificado y no se pueden estimar los parámetros.
- Cero. El número de datos es igual que el número de parámetros estimados, es un modelo saturado y la solución es única. Son modelos no interesantes porque no se simplifica la realidad.

- Positivos. El número de ecuaciones es mayor que el número de incógnitas, el modelo está sobre-identificado y no es posible encontrar un conjunto de variables para los parámetros estimados, pero existe un conjunto de parámetros estimados que hacen que los residuos tomen el valor más pequeño posible para el modelo.

La condición necesaria para poder estimar los parámetros es que el modelo esté sobre-identificado, aunque no es condición suficiente porque pueden existir algunos parámetros que no se pueden estimar. En este caso se dice que el modelo no está identificado localmente.

ESTIMACIÓN DE LOS PARÁMETROS

El objetivo del AFC es conseguir estimaciones de cada uno de los parámetros del modelo que configuren un matriz (S^*) que calque lo más exacto posible la matriz observada (S). Para obtener la estimación de los parámetros hay que buscar los parámetros que minimizan la Función de Discrepancias entre S y S^* , $F(S, S^*)$. Se considera el siguiente vector de discrepancias o residuos:

$$d = \{d_1, \dots, d_l, \dots, d_L\} \quad (17)$$

Donde $d_1 = res_{11}$, $d_2 = res_{12}$, ..., $d_L = res_{JJ}$, siendo L el número de elementos del vector, que es $(J(J + 1))/2$. Esta función toma valor 0 cuando los valores de d son cero y tomará valores mayores cuanto mayor sean los valores de d , en valor absoluto. Existen varios métodos de estimación y los más utilizados son:

- Mínimos Cuadrados no Ponderados (Unweighted Least Squares – ULS). En este método la función de discrepancia es:

$$F_{ULS}(S, S^*) = \sum_l^L d_l^2 \quad (18)$$

Las discrepancias se elevan al cuadrado para que, al minimizar la función, se minimicen dichas discrepancias en valor absoluto. Es el método más sencillo y es adecuado si se analizan las variables estandarizadas.

- Máxima Verosimilitud (ML). La función de discrepancia es:

$$F_{ML}(S, S^*) = \sum_l^L \sum_{l'}^L w_{ll'} d_l d_{l'} \quad (19)$$

En esta función se suman todos los posibles productos entre los elementos del vector de discrepancias ponderando con un peso $w_{ll'}$.

Con este método se buscan los parámetros que hacen más verosímiles las respuestas de los sujetos, asumiendo la distribución de las variables como normal multivariada.

ÍNDICES DE BONDAD DE AJUSTE

Ya estimados los parámetros, el siguiente paso es contrastar el grado en el que el modelo se ajusta a los datos. El número de índices de ajuste para modelos confirmatorios es inmenso, por ello es importante distinguir entre:

- Índices inferenciales. Tienen distribución estadística y permiten hacer inferencias sobre lo que ocurre en la población. Dentro de los índices inferenciales hay dos tipos de ajuste:
 - Ajuste Absoluto. Sirven para establecer en qué grado el modelo reproduce los datos
 - Ajuste Comparativo. Precisan en que grado el modelo propuesto es mejor que otro alternativo.
- Índices descriptivos. Como no tienen distribución estadística conocida, permiten deducir lo que ocurre en la muestra. Dentro de los índices descriptivos también hay ajuste absoluto y comparativo y se definen igual que en los índices anteriores.

Los distintos índices para determinar la bondad de ajuste de un modelo confirmatorio son los siguientes:

- Ji-Cuadrado (χ^2). Corresponde a un índice inferencial de ajuste absoluto y mide la bondad de ajuste global del modelo. El Test del Ji-Cuadrado contrasta la hipótesis nula de que el modelo se ajusta perfectamente a los datos y se acepta esta hipótesis cuando el p-valor es mayor que 0,05. El valor de χ^2 es sensible al tamaño de la muestra y si se acerca a cero, el modelo se ajusta a los datos, cuanto mayor sea el valor, peor es la calidad del ajuste.
- Comparación de modelos anidados. Al utilizar modelos confirmatorios es frecuente que se contrasten distintos modelos a la vez. Puede haber modelos anidados, por ejemplo, uno con dos factores correlacionados y otro con ambos factores independientes, el segundo es el primero, pero con el valor de correlación entre factores igualado a cero. Cuando los modelos son anidados se pueden establecer comparaciones estadísticas del ajuste.
- Residuo Cuadrático Medio (RMR o RMSR). Se obtiene a partir de la Raíz del Promedio de los Residuos al Cuadrado (*Root of Mean Squared Residual*) entre la matriz observada y la estimada. Se puede considerar como aceptable cuando su valor está entre 0,05 y 0,08 pero no tiene fijado un umbral para interpretar su valor.

- Error Cuadrático Medio de Aproximación de la Raíz (RMSEA). Explica la cantidad de varianza no explicada por el modelo por grado de libertad y al tener en cuenta los grados de libertad, es susceptible al número de parámetros que estima el modelo. Los valores de RMSEA varía entre 0 y 1 y decrecen acorde aumenta el número de grados de libertad o el tamaño de la muestra. Se considera un buen ajuste cuando estos valores son inferiores a 0,05 y un ajuste razonable si se encuentran entre 0,05 y 0,08.
- Índice de Bondad de Ajuste (GFI o AGFI). Es una medida de la cantidad relativa de las varianzas y covarianzas en S, pronosticadas por el modelo. Puede tomar valores entre 0 y 1, siendo recomendables los superiores a 0,9 y más cercanos a 1.
- Índice de Ajuste Normalizado (NFI). Analiza la diferencia entre el valor de χ^2 del modelo propuesto y el valor de χ^2 del modelo nulo y no tiene en cuenta los grados de libertad. Tiende a tener un sesgo negativo. Un modelo se considera aceptable si el valor NFI está entre 0,9 y 0,95.
- Índice de Ajuste No Normalizado (NNFI o TLI). Sí tiene en cuenta los grados de libertad porque divide el valor de χ^2 por los grados de libertad del modelo. Se conoce también como el índice de ajuste no normal penalizado porque penaliza los modelos más complejos por agregar parámetros adicionales. También se llama TLI por los autores que lo propusieron (Tucker – Lewis).
- Índice de No Centralidad Relativa (RNI) e Índice de Ajuste Comparativo (CFI). El índice CFI está basado en la comparación del error de aproximación del modelo propuesto y del modelo de independencia. A partir de esos errores de aproximación se calcula el valor RNI

Los índices de bondad de ajuste nombrados se encuentran resumidos en la *Tabla 3*.

ÍNDICES	AJUSTE	ÍNDICE DE AJUSTE	CRITERIO INTERPRETACIÓN
Inferencial	Absoluto	Estadístico χ^2	Se mantiene modelo si $p \geq 0.05$
	Comparativo	Comparación modelos anidados (χ^2)	Modelos no difieren si $p \geq 0,05$
Descriptivo	Absoluto	$\frac{\chi^2}{gl}$	Ratios menores que 2 o 3, indicadores de ajuste aceptable
		RMR	Más pequeño, mejor ajuste. 0 es ajuste perfecto
		SRMR	$\leq 0,08$, el modelo se ajusta
		RMSEA	$\leq 0,06$, el modelo se ajusta
		GFI / AGFI	$\geq 0,95$, el modelo se ajusta
	Comparativo	NFI	$\geq 0,95$, el modelo se ajusta. Tiende a sobrestimar en muestras pequeñas
		NNFI y TLI	$\geq 0,95$ para variables continuas o $\geq 0,96$ para variables categóricas, el modelo se ajusta.
		RNI y CFI	$\geq 0,95$, el modelo se ajusta

Tabla 3. Clasificación de los índices de ajuste

4.2.3 ANÁLISIS DE CORRESPONDENCIAS (ACC)

El Análisis de Correspondencias es una técnica estadística multivariante de análisis de datos categóricos que se utiliza para descubrir y representar estructuras subyacentes en un conjunto de datos. Permite resumir grandes cantidades de información en un número reducido de dimensiones o factores.

- El Análisis de Correspondencias Simples (ACS) es el análisis en el cual los datos tienen forma de tabla de contingencia y se evalúan las relaciones existentes entre dos variables.
- El Análisis de Correspondencias Múltiples (ACM) es el análisis en el cual se trabaja con dos o más variables nominales.

El Análisis de Correspondencias se encuentra dentro de los métodos estadísticos factoriales, diseñado para estudiar las relaciones que hay entre cualquier número de modalidades de las distintas variables. Aquí presenta la diferencia con el Análisis de Componentes Principales, el cual estudia la relación entre las variables y no entre los valores de estas. Es una técnica que se utiliza para representar las filas y columnas de una tabla de contingencia como puntos en un espacio vectorial de baja dimensión, es decir, su principal objetivo es transformar una tabla con información numérica en una representación gráfica que facilite su interpretación.

Este análisis reduce los datos en dos o tres dimensiones, en las cuales sitúa las categorías de las variables que se analizan y los individuos que intervienen en el análisis. Para esto estudia las relaciones entre las categorías de Q variables cualitativas a partir de una muestra con n individuos. En el caso de que exista alguna variable cuantitativa, se puede convertir en cualitativa dividiendo su rango en intervalos de la misma amplitud y cada intervalo será una categoría de la variable.

El Análisis de Correspondencias tiene dos objetivos básicos:

- Asociación entre categorías de columnas o filas. Mide la asociación únicamente de una fila o columna para ver si las modalidades de una variable se pueden combinar.
- Asociación entre categorías de filas y columnas. Estudia si existe relación entre las categorías de las variables de las filas y las columnas.

		Y						
		y_1	y_2	...	y_j	...		y_m
X	x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1m}	$N_{1.}$
	x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2m}	$N_{2.}$

	x_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{im}	$N_{i.}$

	x_k	n_{k1}	n_{k2}	...	n_{kj}	...	n_{km}	$N_{k.}$
		$N_{.1}$	$N_{.2}$...	$N_{.j}$...	$N_{.m}$	$N_{..}$

Tabla 4. Tabla de frecuencias

En la *Tabla 4* se observa la tabla de frecuencias inicial, el punto de partida del ACC, donde la fila de $N_{.m}$ corresponde a los totales marginales columnas, $N_{k.}$ a los totales marginales filas y $N_{..}$ al tamaño total de la muestra.

Si la tabla de contingencia consta de dos variables cualitativas, el objetivo del AC es resumir la información presente tanto en filas como en columnas de manera que sea proyectada sobre un subespacio reducido y representados simultáneamente los puntos fila y puntos columna, obteniéndose así conclusiones entre las dos variables.

CONCEPTOS IMPORTANTES

Un *perfil* es la distribución de frecuencias relativas de una línea de la tabla, ya sea fila o columna, en relación con su total marginal. El perfil i -ésimo es:

$$\left(\frac{n_{i1}}{N_{i.}}, \frac{n_{i2}}{N_{i.}}, \dots, \frac{n_{ij}}{N_{i.}} \right) \quad (20)$$

Los perfiles columna describen las distribuciones condicionadas de la variable Y por las distintas modales de la variable X y los perfiles fila describen las distribuciones condicionadas de la variable X por las distintas modalidades de la variable Y.

Una *masa* es el cociente entre el total marginal de la línea, sea fila o columna y el total marginal que es asignada a cada perfil fila o columna. Es decir, la distribución marginal de la línea con respecto al total global. Por ejemplo, la masa del i -ésimo perfil fila es

$$r_i = \frac{N_{i.}}{N_{..}} \quad (21)$$

El uso de perfiles facilita la interpretación, pero puede producir una visión equivocada de la relación que hay entre variables. Para solucionar este problema se utiliza una distancia que no ignore las diferencias entre los perfiles de cada línea. La *distancia ji-cuadrado* es una

distancia euclídea ponderada que sirve para medir distancias entre perfiles. Se calcula de forma que las categorías que son más frecuentes se ponderan menos y las que son menos frecuentes se ponderan más.

Dados dos puntos $P_i = (X_{i1}, X_{i2})$ y $P_{i'} = (X_{i'1}, X_{i'2})$ la distancia euclídea entre ellos es

$$d^2(P_i, P_{i'}) = (x_{i1} - x_{i'1})^2 + (x_{i2} - x_{i'2})^2 \quad (22)$$

Por tanto, dados dos perfiles

$$i : \left(\frac{n_{i1}}{N_{i.}}, \frac{n_{i2}}{N_{i.}}, \dots, \frac{n_{ij}}{N_{i.}} \right) \quad (23)$$

$$i' : \left(\frac{n_{i'1}}{N_{i'.}}, \frac{n_{i'2}}{N_{i'.}}, \dots, \frac{n_{i'j}}{N_{i'.}} \right)$$

La distancia ji-cuadrado entre ellos será:

$$d^2(i, i') = \sum_{j=1}^p \left(\frac{1}{N_{.j}} \right) \left(\frac{n_{ij}}{N_{i.}} - \frac{n_{i'j}}{N_{i'.}} \right) \quad (24)$$

La distancia ji-cuadrado posee la propiedad de equivalencia distribucional. Esta propiedad garantiza que en espacios dotados de distancia ji-cuadrado se pueden reemplazar dos filas (o columnas) sin que esto modifique las distancias existentes entre las columnas (o filas). Es decir, si dos categorías de los perfiles fila tienen el mismo valor de perfil, al agruparlas en una sola categoría, no se modifican las distancias existentes entre el resto de las categorías. No hay pérdida de información al reunir clases con perfiles idénticos.

Las distancias no se miden entre filas o columnas, sino con relación a un perfil medio de fila o columna, es decir, con relación al promedio de las coordenadas de dicha línea ponderada por su masa. Este perfil medio se sitúa en el origen de coordenadas y se denomina *centro de gravedad* o *centroide*.

La *inercia* de una nube de puntos, la cual representa los perfiles, es una medida de la dispersión de los perfiles en un espacio multidimensional. Se calcula como la suma, para cada punto, de su masa por su distancia al centroide elevada al cuadrado. La inercia total de una tabla coincide con el valor del ji-cuadrado correspondiente. Si el valor de la inercia es bajo, significa que todos los productos están situados cerca del centroide y que, por tanto, son similares, mientras que, valores altos de inercia, indican lo contrario. El Análisis de Correspondencias trata de encontrar planos en los que la inercia sea máxima.

Otro concepto importante en el ACC son los *valores propios*, que representan la inercia de la dimensión correspondiente y se pueden interpretar como la correlación entre las puntuaciones columna y fila. Los valores propios se representan en un gráfico en el que las alturas son proporcionales a sus valores, permitiendo evaluar el crecimiento de los valores propios. Que uno de ellos esté separado del anterior y del siguiente, es una señal de que el

eje correspondiente está bien separado y por ello será razonable buscar su interpretación. Si dos valores propios se encuentran muy juntos pero separados del resto, indica que sus planos correspondientes están mejor individualizados que sus ejes.

La proyección de un perfil, el cual representa a una categoría, sobre una de las dimensiones, da lugar a una coordenada que se llama *puntuación de la categoría en la dimensión*.

Una vez explicados los conceptos básicos del ACC, el análisis procede a diagonalizar la matriz de varianza – covarianza (C) con el fin de obtener los vectores y valores propios que van a definir los nuevos ejes sobre los que se proyectará la nube de puntos.

El ACC intenta encontrar dos matrices llamadas A y B de coordenadas cartesianas

$$A = \begin{pmatrix} a'_{11} \\ a'_{12} \\ \dots \\ a'_{1k} \end{pmatrix}, \quad B = \begin{pmatrix} b'_{11} \\ b'_{12} \\ \dots \\ b'_{1k} \end{pmatrix} \quad (25)$$

donde A representa a los puntos fila y B los puntos columna. Hay distintos métodos de calcular estas matrices y se llaman *normalizaciones*:

- Simétrico. En cada dimensión las puntuaciones de fila son el promedio ponderado de las puntuaciones de columna entre el autovalor correspondiente. Este método trata las filas y columnas de forma simétrica y, por tanto, distribuye la inercia de manera idéntica sobre las coordenadas de las filas y columnas. Es el método más utilizado si el objetivo es describir los patrones de asociación entre variables.
- Principal por fila. Las distancias euclídeas entre los puntos fila aproximan la distancia ji-cuadrado existente entre las filas de la tabla. Se realiza cuando el objetivo es ver como difieren las categorías de las variable fila entre sí. Existe el mismo método, pero por columna.
- Principal. Este método permite examinar las diferencias o semejanzas entre las categorías de variables, pero por separado. Se utiliza cuando el objetivo es explicar las distancias existentes entre puntos fila y entre puntos columnas de manera independiente.
- Personalizado. Se especifica un valor incluido en el intervalo [-1, 1], donde más próximo a 1 es igual al método principal por objeto y más próximo a -1 es igual al método principal por fila/columna. Es un método que se utiliza cuando el objetivo es crear diagramas de dispersión biespaciales y triespaciales a medida.

INTERPRETACIÓN

- La *inercia total* (IT) mide el grado de dependencia que hay entre las variables.
- Las *proporciones de inercia explicada* se calculan a partir de la inercia total y ayudan a ponderar la importancia que tiene cada una de las dimensiones.
- Las *proporciones de inercia acumulada explicada* por las *i*-ésimas primeras dimensiones dan lugar a la decisión del número mínimo de dimensiones que son necesarias para explicar las dependencias.
- Las *contribuciones totales* miden la importancia que tiene cada modalidad de la variable analizada en la construcción de los ejes. Se utilizan para interpretar el significado que tiene cada eje utilizando en cada uno de ellos la modalidad con contribución más fuerte.
- Las *contribuciones relativas* miden la importancia que tiene cada factor para explicar la posición de cada modalidad. Se utilizan para analizar las proximidades que hay entre los puntos, teniendo en cuenta los factores con contribuciones más altas.
- La proximidad entre individuos indica que los individuos se parecen, que tienen las mismas modalidades.
- Si dos categorías aparecen como puntos muy próximos en el gráfico, quiere decir que presentan asociación positiva.

5. RESULTADOS

5.1 ANÁLISIS DESCRIPTIVO DE LOS DATOS

La muestra está formada mayoritariamente por mujeres (68,7%) y una media de edad comprendida entre 26 y 45 años (39,6%). La muestra fue recogida en el territorio español y más de la mitad corresponde a individuos residentes en Salamanca (55,5%), seguido de Madrid (7,5%) y de Barcelona (4,6%), en el *Gráfico 1* se pueden observar las provincias de las cuales se han obtenido datos. Si la muestra de cada provincia hubiera sido más representativa, se habría pensado en hacer algún análisis teniendo en cuenta estos datos, pero si se realiza, los resultados obtenidos no serían válidos porque la variable que corresponde con la provincia de residencia estaría desbalanceada.



Gráfico 1. Provincias de la muestra

En el *Gráfico 2* se comprueba que la comunidad autónoma de la cual se han obtenido más datos es Castilla y León debido a la cantidad de casos obtenidos de Salamanca (63%).

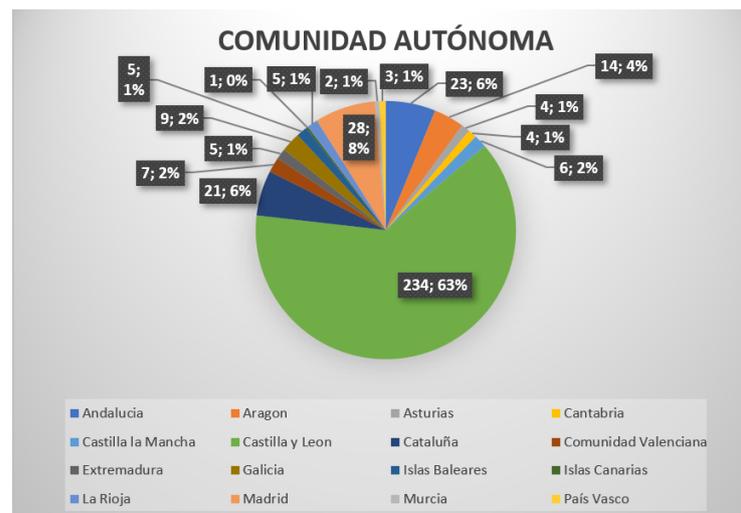


Gráfico 2. Comunidades autónomas de la muestra

En cuanto a la zona de residencia, la mayoría de los individuos residen en una zona urbana (77,6%), algo que tiene sentido porque en las zonas rurales es más rara la presencia de supermercados. En relación con la compañía en el domicilio, el grupo más grande es el de individuos que viven con su pareja (36,7%).

Respecto a la situación laboral de los individuos encuestados, más de la mitad son trabajadores (58,8%). Un dato interesante es que el porcentaje de los individuos desempleados sin prestación (7,5%) sumado al porcentaje de los individuos que se encuentran en paro o ERTE (9,5%) alcanza casi una quinta parte de la muestra, algo que seguramente sea debido a la crisis producida por la pandemia vivida.

5.2. ANÁLISIS DE LA ESTRUCTURA FACTORIAL DEL CUESTIONARIO

Al realizar el Alfa de Cronbach se observa que hay algún ítem invertido, pero tras recodificarlos correctamente se obtiene un valor de $\alpha = 0,718$.

Los valores del determinante de la matriz de correlaciones (0,009), la medida de Kaiser-Meyer-Olkin (0,705) y la prueba de esfericidad de Bartlett (p-valor < 0,05) indican que es apropiado realizar un Análisis Factorial de los datos.

Se ha realizado un Análisis Factorial Exploratorio en el que se han incluido las 21 variables que corresponden a las no sociodemográficas.

Según la regla de Kaiser se identifican ocho factores que tienen un autovalor mayor que 1, los cuales explican casi el 60% de la varianza de respuesta del test. En la interpretación del gráfico de sedimentación (Scree Plot) que se muestra en el *Gráfico 3*, se pueden considerar tres factores antes de que el gráfico empiece a tender a una línea recta.

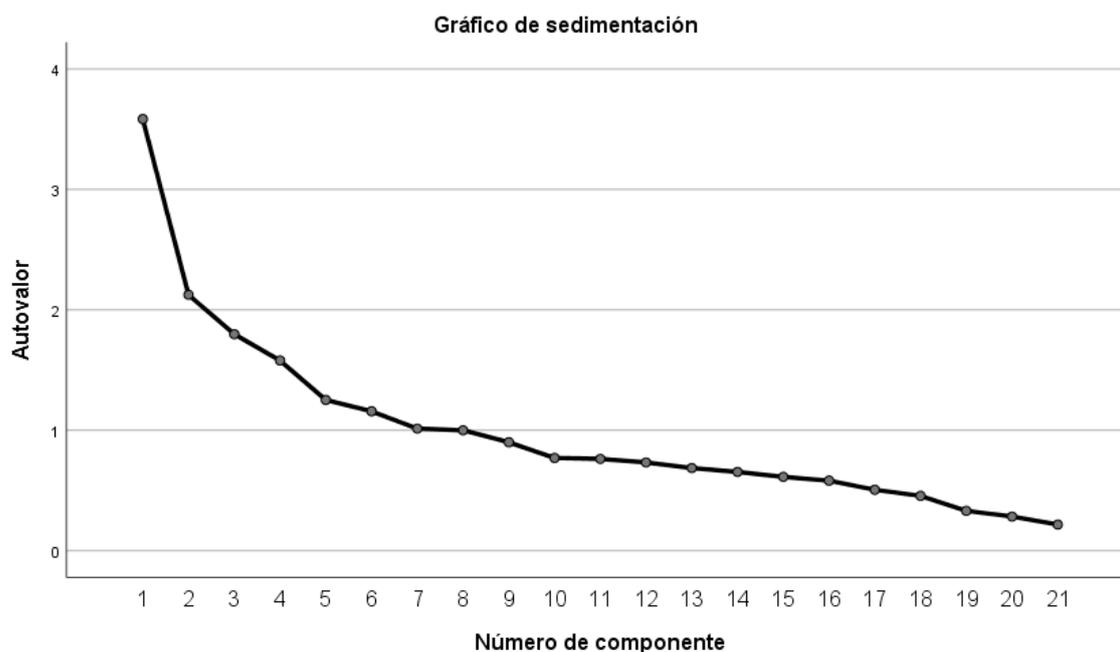


Gráfico 3. Gráfico de sedimentación (path diagram)

Usando 7 factores algunos son incorrectos por explicar únicamente una o dos variables. Se decide, por tanto, realizar el análisis con tres factores, los cuales explican un 36% de la varianza por considerar que pueden explicar bien los datos y que la agrupación de las variables tiene sentido.

En la *Tabla 5* se encuentran los pesos factoriales rotados mediante el método *Varimax* que tiene cada variable en cada factor. Se considera que una variable está explicada por el factor que mayor peso factorial tiene en valor absoluto.

	COMPONENTE		
	1	2	3
COMPANT	0,080	0,205	0,048
NUMSUP	0,001	-0,076	-0,031
CAMBIO	0,181	0,410	0,314
NUMDES	0,164	0,474	-0,010
FREQCOM	0,009	0,602	0,091
TIEMCOMP	-0,036	0,657	-0,085
PLAN	-0,031	0,560	-0,066
CALIDAD	0,062	0,006	0,424
GASTO	-0,032	0,063	0,744
MOTAUM	-0,011	0,270	0,800
TIRCOM	0,087	-0,148	0,452
GASTDES	0,033	0,325	0,422
ONLSUP	0,601	-0,090	0,145
ONLINE	0,342	0,125	0,220
PQONLSUP	0,766	0,220	0,061
ONLDES	0,695	0,003	0,072
MESONL	0,830	0,043	0,034
INTONL	0,599	0,034	0,127
COMPRES	0,106	0,558	-0,271
COMPONL	0,581	0,130	-0,100
PRESO NL	0,499	0,051	-0,130
α	0,751	0,560	0,724

Tabla 5. Matriz de componente rotado (Varimax). COMPANT: Lugar de compra antes de la pandemia. NUMSUP: Compra en un sup. o varios. CAMBIO: Cambio en los hábitos de compra. NUMDES: Número de compras al mes desde el comienzo de la pandemia. FREQCOM: Motivo del cambio de la frecuencia de compra. TIEMCOMP: Dedicar el mismo tiempo que antes de la pandemia a la compra. PLAN: Planificar la compra. CALIDAD: Calidad de los productos comprados. GASTO: Variedad en el gasto de la compra. MOTAUM: Motivos para aumentar el gasto en compra. TIRCOM: Comida tirada tras la pandemia. GASTDES: Cambio en los gastos cuando acabe la pandemia. ONLSUP: Compra online en sup. antes de la pandemia. ONLINE: Compra online antes de la pandemia. PQONLSUP: ¿Por qué compras online en sup. antes de la pandemia? ONLDES: Compra online en sup. desde que empezó la pandemia. MESONL: Compras online al mes en sup. desde que empezó la pandemia. INTONL: Interesante la compra online. COMPRES: Compra presencial durante el confinamiento comparada con antes. COMPONL: Compra online durante el confinamiento comparada con antes. PRESO NL: Sustituir compra presencial por online.

En la teoría inicial del análisis se pensó en cuatro factores llamados “Supermercado”, “Compra”, “Economía” y “Online” pero tras realizar el Análisis Factorial Exploratorio con tres factores se unen los ítems que corresponden a los dos primeros factores teóricos.

El primer factor designado “Compra Online” está integrado por los ítems ONLSUP (compra online en sup. antes de la pandemia), ONLINE (compra online antes de la pandemia), PQONLSUP (¿por qué compras online en sup. antes de la pandemia?), ONLDES (compra online en sup. desde que empezó la pandemia), MESONL (compras online al mes en sup. desde que empezó la pandemia), INTONL (Interesante la compra online), COMPONL (compra online durante el confinamiento comparada con antes), PRESONL (sustituir compra presencial por online).

El segundo factor llamado “Compra” está compuesto por los ítems COMPANT (lugar de compra antes de la pandemia), NUMSUP (compra en un sup. o varios), CAMBIO (cambio en los hábitos de compra), NUMDES (número de compras al mes desde el comienzo de la pandemia), FREQCOM (motivo del cambio de la frecuencia de compra), TIEMPCOMP (dedicar el mismo tiempo que antes de la pandemia a la compra), PLAN (planificar la compra) y COMPRES (compra presencial durante el confinamiento comparada con antes). Este último ítem debería pertenecer al primer factor, pero con los datos obtenidos el análisis lo incluye en este. El motivo puede ser que, como ese ítem pregunta sobre la compra presencial, lo relacione con el factor correspondiente a la compra y el supermercado.

El tercer factor, denominado como “Economía” está integrado por los ítems CALIDAD (calidad de los productos comprados), GASTO (variedad en el gasto de la compra), MOTAUM (motivos para aumentar el gasto en compra), TIRCOM (comida tirada tras la pandemia) y GASTDES (cambio en los gastos cuando acabe la pandemia).

La consistencia interna de cada factor se verifica de nuevo mediante el coeficiente Alfa de Cronbach para el cual se consigue un valor aceptable para el primer factor (0,751) y algo menos para el factor dos (0,56) y de nuevo aceptable para el tercer factor (0,724). Como se indicó al principio de los resultados, el valor de α para la prueba total es 0,718, lo que indica una buena consistencia interna y fiabilidad en la prueba general.

Una vez finalizado el Análisis Factorial Exploratorio se ha realizado un Análisis Factorial Confirmatorio, obteniendo unos resultados que indican que las dimensiones propuestas por el AFE son aceptables para los datos.

Según Thompson (2004), al emplear el Análisis Factorial Confirmatorio, no solo debe aprobarse el ajuste del modelo teórico, sino que es recomendable comprar los índices de ajuste de distintos modelos alternativos para seleccionar el mejor. Para elegir el modelo que se puede observar en el *Gráfico 4*, se comparó dicho modelo junto a otro con 4 factores (el modelo teórico).

En la *Tabla 6* se encuentran los índices de bondad de ajuste del modelo esperados y los que se han obtenido en el modelo con cuatro y con tres factores, seleccionando como mejor el modelo con tres factores. En ambos casos el estadístico χ^2 resulta significativo (p-valor < 0,05), lo cual es debido a la sensibilidad que presenta al tamaño de la muestra. Por esto, se

proponen otros índices de ajuste como son el Índice de Bondad de ajuste (GFI = 0,870), el índice Ajustado de Bondad de Ajuste (AGFI = 0,879), el Índice Residual de la Raíz Cuadrática Media (RMR = 0,05), el Error Cuadrático Medio de Aproximación (RMSEA = 0,058), el Índice de Ajuste Comparativo (CFI = 0,885), el Índice de Ajuste Normalizado (NFI = 0,790) y el Índice No Normalizado de Ajuste (NNFI o TLI = 0,383).

Los resultados del Análisis Factorial Confirmatorio muestran que los índices no son totalmente aceptables, pero como se acercan mucho al punto óptimo, se pueden considerar buenos, es decir, el modelo se ajusta bien a los datos.

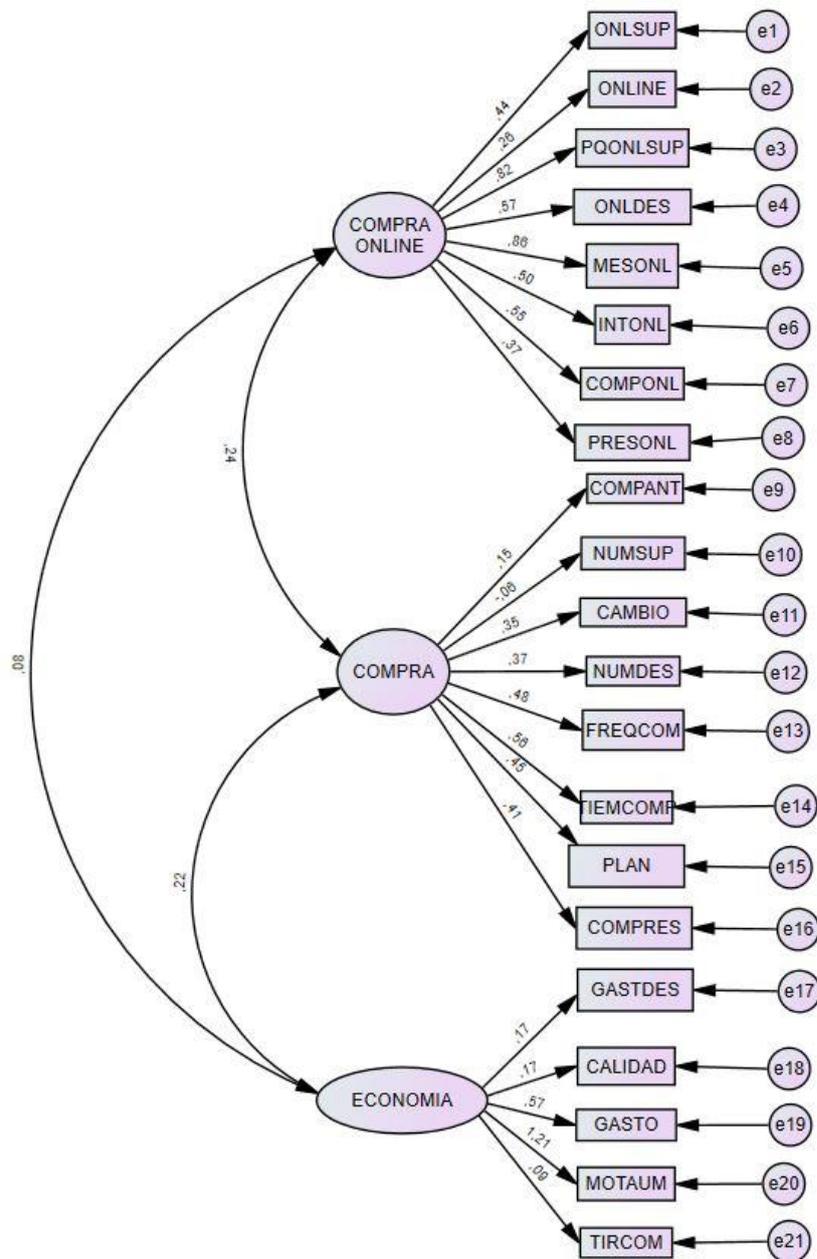


Gráfico 4. Gráfico del Análisis Factorial Confirmatorio

ÍNDICE DE AJUSTE	ESPERADO	3 FACTORES	4 FACTORES
P – valor χ^2	> 0,05	0,000	0,000
Discrepancia entre χ^2 y los grados de libertad (CMIN/DF)	< 5	2,684	3,223
Índice de Bondad de Ajuste (GFI)	0,9 – 1	0,870	0,886
Índice de Ajuste Ponderado (AGFI)	0,9 – 1	0,879	0,855
Índice Residual de la Raíz Cuadrática Media (RMR)	Próximo 0	0,05	0,058
Error Cuadrático Medio de Aproximación (RMSEA)	< 0,05	0,058	0,078
Índice de Ajuste Comparativo (CFI)	0,9 – 1	0,885	0,734
Índice de ajuste Normalizado (NFI)	0,9 – 1	0,790	0,660
Índice No Normalizado de Ajuste (TLI)	0,9 - 1	0,838	0,694

Tabla 6. Índices de Bondad de Ajuste del AFC

5.3. PERFIL DE LOS CONSUMIDORES DURANTE LA COVID-19

Para crear unos perfiles y caracterizar a los individuos, se utiliza la variable “cambio”, la cual tiene en cuenta si cada individuo encuestado considera que han cambiado sus hábitos en los supermercados tras la Covid-19. Las variables con las que se realiza el Análisis son las variables que están relacionadas con “cambio”, es decir, que tras realizar la prueba Ji-Cuadrado, resultan significativas.

Se incluyen las tablas y gráficos de la relación con la primera variable y el resto serán incluidos en el *Anexo II*.

La primera variable con la que se hace el análisis es la variable “edad”.

CAMBIO	EDAD				
	18-25	26-45	46-65	Mayor de 65	Margen activo
Si, se han visto afectados a peor	22	31	10	2	65
Si, se han visto afectados a mejor	21	25	12	0	58
No, se mantienen igual	65	91	84	8	248
Margen activo	108	147	106	10	371

Tabla 7. Tabla de Correspondencias de las variables cambio x edad

En la *Tabla 7* se muestran los valores correspondientes a las combinaciones entre las categorías de las variables Cambio (X) y Edad (Y) y los totales marginales de fila y columna.

Dimensión	Valor singular	Inercia	Chi cuadrado	Sig.	Proporción de inercia		Valor singular de confianza	
					Contabilizado para	Acumulado	Desviación estándar	Correlación 2
1	0,180	0,032			0,885	0,885	0,047	-0,202
2	0,065	0,004			0,115	1,000	0,037	
Total		0,036	13,516	0,036	1,000	1,000		

Tabla 8. Tabla del resumen del análisis

En la *Tabla 8* se encuentra el resumen del análisis. El número de dimensiones corresponde con el mínimo entre las categorías de cada variable menos uno, es decir $K = \min\{3-1, 4-1\} = 2$ dimensiones. En esta tabla se muestran las contribuciones de cada una de estas dimensiones a la inercia total. Se puede observar que solo la primera dimensión contribuye un 88,5% a la inercia total, esto significa que las dependencias que se observan en la tabla vienen perfectamente capturadas por la primera dimensión. Las dependencias observadas en la tabla son capturadas adecuadamente por la primera dimensión.

CAMBIO	Masa	Puntuación en dimensión		Inercia	Contribución				
		1	2		Del punto en la inercia de dimensión		De la dimensión en la inercia del punto		Total
					1	2	1	2	
Si, a peor	0,175	-0,649	0,391	0,015	0,411	0,414	0,885	0,115	1
Si, a mejor	0,156	-0,546	-0,492	0,011	0,259	0,585	0,774	0,226	1
No, igual	0,668	0,298	0,013	0,011	0,330	0,002	0,999	0,001	1
Total activo	1			0,036	1	1			

Tabla 9. Contribuciones totales y relativas de los perfiles fila

La *Tabla 9* muestra las contribuciones totales y relativas. Las contribuciones del punto a la inercia de la dimensión es la parte que toma la categoría en la varianza explicada por esa dimensión. Los elementos que tienen mayores contribuciones son los que más contribuyen a la posición de los ejes. En estas contribuciones, la categoría con mayores valores es “No, igual” para la primera dimensión y “Sí, a mejor” para la segunda. Las contribuciones de la dimensión a la inercia del punto es el porcentaje de buena representatividad de las variables. Las categorías que mayor contribución tengan son esenciales para interpretar la dimensión correspondiente. Para estas contribuciones la mayor para la dimensión 1 es “No, igual” y “Sí, a mejor” para la dimensión 2, igual que para el caso anterior.

Otro dato a observar es que, en la parte positiva del primer factor en la puntuación en la dimensión, se encuentran los individuos que no han sufrido cambios, y en la negativa los que sí los han sufrido, esto significa que el primer factor discrimina por el cambio en los hábitos.

La tabla de contribuciones totales y relativas de los perfiles columna es igual, pero con las categorías de la variable "edad". Para las contribuciones del factor al elemento es mayor el grupo de edad 46-65 para la primera dimensión y mayor de 65 para la segunda dimensión. En las contribuciones de la dimensión a la inercia del punto es mayor la categoría de edad entre 46 y 65 años para la primera dimensión y mayores de 65 para la segunda.

El *Gráfico 5* muestra la asociación entre las categorías de ambas variables. Este gráfico pone de manifiesto las relaciones de dependencia que existen entre las dos variables. En el gráfico se observa lo dicho anteriormente, la primera dimensión discrimina entre los individuos que sí y no han sufrido cambios. Para la segunda dimensión se podría decir que discrimina entre mayores y menos de 25 años.

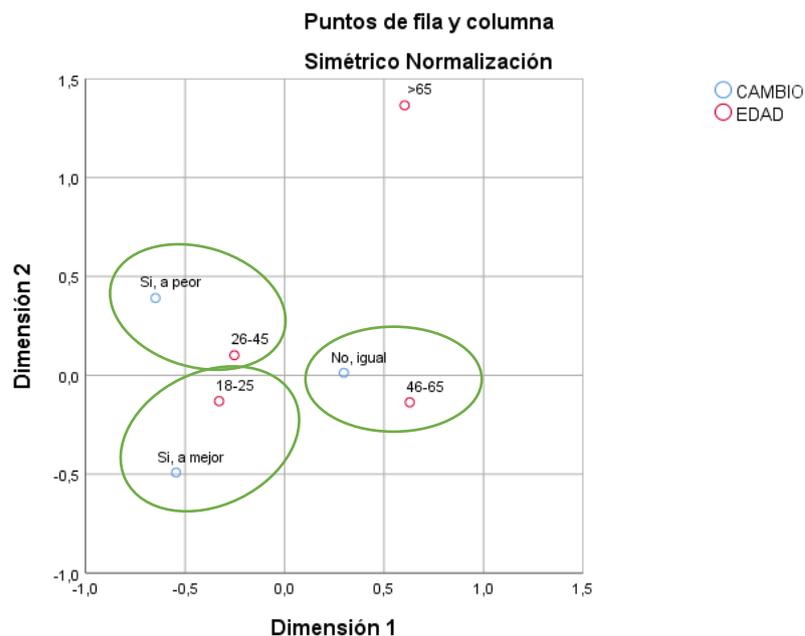


Gráfico 5. Gráfico conjunto de los puntos correspondientes a los perfiles condicionales fila y columna

Con la información que aporta el *Gráfico 5* se obtiene la siguiente información de los consumidores: Los individuos que no han visto modificado su comportamiento en los supermercados durante la Covid-19 son individuos que se encuentran entre los 46 y 65 años de edad. Por lo general, las personas que se encuentran en esas edades tienen trabajos o sueldos fijos, algo que, tras vivir la crisis provocada por la pandemia no les hará ver afectados sus sueldos o sus costumbres. Los consumidores que están entre 18 y 25 años de edad sí han visto modificados sus hábitos, pero a mejor. Esto puede ser porque hayan encontrado trabajo y hayan mejorado sus condiciones. Por último, los individuos que están entre 26 y 45 años han vivido un cambio a peor. Aquí puede ser lo contrario al primer caso, pueden ser individuos que tenían trabajos temporales o que han sido despedidos y por ellos han visto peor su situación tras la pandemia.

Teniendo en cuenta los gráficos obtenidos con el resto de las variables, los cuales se encuentran en el *Anexo II*, los perfiles que se obtienen de los consumidores en los supermercados durante la Covid-19 son los siguientes:

Los consumidores que creen que sus hábitos de consumo en los supermercados se han visto afectados a peor son individuos entre 25 y 45 años que, tras comenzar la pandemia, han visto disminuir sus ingresos entre un 10 y 100% y que realizan la compra cada dos semanas, habiendo disminuido su frecuencia por miedo al contagio y realizando las compras más despacio para tener controlado el gasto, el cual ha aumentado por realizar más comidas en casa y menos en restaurantes o en pedidos a domicilio. Siguiendo con la importancia que le dan al gasto, son individuos que tras la pandemia tienen pensado cambiarlo porque comenzarán a darle más valor a los precios. Respecto a la compra online, la realizaban antes de la pandemia por comodidad y por ser más económico al evitar tentaciones y por este mismo motivo realizan más compra de esta forma tras el comienzo de dicha pandemia, llegando a sustituir la compra presencial por compra online.

Los consumidores que consideran que no han cambiado sus hábitos de consumo en los supermercados son individuos entre 46 y 65 años que, desde que empezó la pandemia no han visto modificados sus ingresos y que realizan la compra una, dos o tres veces por semana, manteniendo igual su frecuencia y sin importar el tiempo dedicado a ella. El gasto que esperan hacer una vez acabe la pandemia es el mismo que estaban teniendo hasta el momento. Respecto a la compra online son consumidores que no la realizaban en supermercados antes de que comenzara la pandemia y tampoco tras su inicio y por tanto no han modificado la compra presencial por online, sino que han mantenido únicamente la primera.

Los consumidores que creen que sus hábitos de consumo en los supermercados se han visto afectados a mejor son individuos entre 18 y 25 años que, tras comenzar la pandemia, han visto aumentar sus ingresos. Han ampliado también la frecuencia de compra, realizándola por lo general todos los días para buscar ofertas/promociones y distintos tipos de marcas o productos y haciéndolas más deprisa para pasar menos tiempo en el supermercado, además de más grandes, con más cantidad de productos. Teniendo en cuenta el gasto en las compras, pretenden que disminuya mediante compras más pequeñas y organizadas. Son individuos que han empezado a comprar online tras el comienzo de la pandemia por miedo al contagio pero que pretenden mantener las compras tanto online como presenciales una vez finalizada la pandemia.

6. CONCLUSIONES

- La consistencia interna del cuestionario se evaluó mediante el coeficiente Alfa de Cronbach y este valor aceptable fue de 0,718.
- La validez del cuestionario sobre el comportamiento de los consumidores durante la Covid-19 en los supermercados se midió con el Análisis Factorial Exploratorio (AFE) y lanzó tres factores a través de la organización de los distintos ítems, siendo aceptable la consistencia interna de cada una de las dimensiones.
- Los resultados obtenidos muestran que la estructura factorial contiene 3 factores
- Como resultado del Análisis Factorial Confirmatorio (AFC) se concluyó que el modelo se ajustaba correctamente a los datos y que era más adecuado el modelo con tres factores que con cuatro.
- Los perfiles de los consumidores se obtienen mediante distintos Análisis Factorial de Correspondencias Simple (ACS), los cuales devuelven tres tipos de consumidores:
 - Los que han vivido un cambio para peor, son individuos en general que han perdido ingresos y que centran sus compras en el dinero.
 - Los que no han vivido cambio, son individuos que mantienen todos sus hábitos igual.
 - Los que han vivido un cambio a mejor, son individuos que no se centran tanto en el dinero, pero si en el miedo y la exposición al virus.

7. BIBLIOGRAFÍA

- Casal, J., & Mateu, E. (2003). Tipos de muestreo. *Rev. Epidem. Med. Prev*, 1(1), 3-7
- Hernández Ávila CE, Carpio N. (2019) Introducción a los tipos de muestreo. *Revista ALERTA*; 2(1): 75-79. DOI: 10.5377/alerta.v2i1.7535
- VII, C. (2000). Análisis factorial
- Ferrando, P. J., & Anguiano-Carrasco, C. (2010). El análisis factorial como técnica de investigación en psicología. *Papeles del psicólogo*, 31(1), 18-33
- Cuadras, C. M. (1996). Nuevos métodos de análisis multivariante. CMC Edicions
- Cortina, J. M. (1993). What is Coefficient Alpha? An Examination of Theory and Applications. *Journal of Applied Psychology*, 78(1), 98-104.
- Pérez, E. R., & Medrano, L. A. (2010). Análisis factorial exploratorio: bases conceptuales y metodológicas. *Revista Argentina de Ciencias del Comportamiento (RACC)*, 2(1), 58-66
- De la Fuente Fernández, S. (2011). Análisis factorial. *Univ Auton Madrid*, 1-34.
- “Análisis Factorial”. *Métodos Estadísticos para Economía y gestión*.
- Mavrou, I. (2015). Análisis factorial exploratorio: cuestiones conceptuales y metodológicas. *Revista Nebrija de Lingüística Aplicada a la Enseñanza de Lenguas*, (19).
- Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical assessment, research, and evaluation*, 10(1), 7.
- Frías-Navarro, D., & Soler, M. P. (2013). Prácticas del análisis factorial exploratorio (AFE) en la investigación sobre conducta del consumidor y marketing. Artículos en PDF disponibles desde 1994 hasta 2013. A partir de 2014 visítenos en www.elsevier.es/sumapsicol, 19(1).
- Abad, F. J. (2011). *Medición en ciencias sociales y de la salud. Síntesis*.
- Morata-Ramírez, M., Holgado-Tello, F. P., Barbero-García, I., & Mendez, G. (2015). Análisis factorial confirmatorio: recomendaciones sobre mínimos cuadrados no ponderados en función del error Tipo I de Ji-Cuadrado y RMSEA. *Acción psicológica*, 12(1), 79-90.
- De Frutos, B., Ruiz, M., & San Martín, R. (1998). Análisis factorial confirmatorio de las dimensiones del compromiso con la organización. *Psicológica*, 19(3), 345-366.

- Santisteban Requena, C. (2009). Principios de psicometría. Madrid: Síntesis.
- Frías-Navarro, D. (2014). Apuntes de SPSS. Universidad de Valencia, 1-10.
- Ruiz Bueno, A. (2015). Validez y fiabilidad. Conceptualización y procedimientos de cálculo con Spss.
- López-Aguado, M., & Gutiérrez-Provecho, L. (2019). Com dur a terme i interpretar una anàlisi factorial exploratòria utilitzant SPSS. *REIRE Revista d'Innovació i Recerca en Educació*, 12(2), 1-14.
- Camarero Sierra, Á. L. (2017). ANÁLISIS FACTORIAL CONFIRMATORIO: APLICACIÓN AL CÁLCULO DE ÍNDICES EN ECONOMÍA.
- Olivares, J. E. P. (1996). Modelo de análisis de correspondencias múltiples. *Revista de Ciencias Sociales*, 2(2), 183-196.
- De la Fuente, S. (2011). *Análisis Correspondencias Simples y Múltiples* [Universidad Autónoma de Madrid].
- Fernandez, J. (2021). Investigador, infiel y polarizado. Así es el consumidor tras un año de pandemia. [online] El País. Available at: <<https://elpais.com/economia/estar-donde-estes/2021-03-15/investigador-infiel-y-polarizado-asi-es-el-consumidor-tras-un-ano-de-pandemia.html>> [Accessed 3 July 2021].
- Guach, R. A. D., Tejeda, J. J. G., & Abreu, M. R. P. (2020). Características clínico-epidemiológicas de la COVID-19. *Revista Habanera de Ciencias Médicas*, 19(2), 1-15.
- Batista-Foguet, J. M., Coenders, G., & Alonso, J. (2004). Análisis factorial confirmatorio. Su utilidad en la validación de cuestionarios relacionados con la salud. *Medicina clínica*, 122(1), 21-27.
- Bueno, J. (2020). Análisis del comportamiento del shopper durante la Covid-19. [online] Food Retail. Available at: <https://www.foodretail.es/blogs/firma_invitada/Analisis-comportamiento-shopper-Covid-19_7_1460323955.html> [Accessed 6 July 2021].
- Da Silva, E. (2020). ¿Cuál es el comportamiento del consumidor después del COVID-19?. [online] lemondigital. Available at: <<https://lemon.digital/comportamiento-del-consumidor-despues-de-pandemia-covid-19/>> [Accessed 6 July 2021].
- Rigoberto Casco, A. (2020). INNOVARE Revista de Ciencia y Tecnología. Efectos de la pandemia de COVID-19 en el comportamiento del consumidor.

8. ABSTRACT

8.1. INTRODUCTION

At the beginning of 2020, a virus originated in Wuhan (China) called SARS-CoV_2 arrived in Spain which has caused more than 80 thousand deaths in Spain until July 2021. The pandemic that this disease has produced has caused many changes in the lives of the Spanish inhabitants. After the state of alarm caused by the Spanish government, mobility was restricted with the except for working or purchasing basic products, which led to a major economic crisis.

Crisis, restrictions and fear led to many families to change their consumption habits. Big and less frequent purchases, regular supermarket change, price-based brand changes, more online purchases, and many more changes.

The main objective of the work is to know how Spanish consumers behave in supermarkets after the beginning of the pandemic, from a multivariate perspective.

8.2. OBJETIVES

Measure the reliability and analyze the factorial structure of the questionnaire used to measure consumer behavior in supermarkets.

Characterize consumers and find out how they behave according to whether their habits have been modified by the current pandemic.

8.3. THEORETICAL FRAMEWORK

On 14 March 2020, the Council of Ministers approved the declaration of the state of alarm throughout Spain for 15 calendar days with measures restricting people's movement, allowing them to leave their homes only for essential activities (work, shopping, dependent care, etc.).

The appearance of the virus caused a major crisis due to the state of alarm that forced the closure of businesses, the restriction of mobility, online studies and the adaptation to teleworking. This crisis left many families without money to make ends meet.

Due to the fear produced by the virus, purchases were made with the objective of filling the houses, without taking into account the quantities, which were often disproportionate. One of the biggest changes produced by the pandemic is the massive use of electronic devices.

8.4. METHODOLOGY

8.4.1. DATA DESCRIPTION

The data for the analysis are obtained from a questionnaire on the consumption of individuals in supermarkets during Covid'19. This questionnaire consists of 29 categorical questions, of which eight are socio-demographic questions and the rest are divided into four

parts which are supermarket, shopping, economy and online. Each of them is intended to measure issues related to the name of the dimension.

The sample was taken over a period of approximately 2 weeks and the target population of the study was the population over 18 years of age in the Spanish region who buy in at least one supermarket. Due to the current situation, the survey was conducted using *Google Forms* in order to avoid physical contact and to allow the access to more individuals, which biases the sample because the same number of data is not obtained from all age groups. Despite this, information is obtained on 271 individuals from 39 provinces.

The sampling method used was a mixed sampling method using purposive or convenience sampling and snowball sampling.

IBM SPSS Statistics 22 and IBM SPSS Amos 26 were used to carry out the respective analyses.

8.4.2. RELIABILITY

The reliability of a test refers to its precision when used as a measuring instrument. Reliability can be understood in three different ways: intra-reliability, test-retest reliability and inter-reliability.

8.4.3. FACTOR ANALYSIS (FA)

Factor analysis is a multivariate statistical technique that represents the relationships between a set of variables and allows to reduce the size of the data without losing too much information.

EXPLORATORY FACTOR ANALYSIS (EFA)

Exploratory Factor Analysis is a set of multivariate statistical methods of interdependence (all variables are analysed together, there is no independent variable and no response variables) that aim to identify a structure of factors underlying a set of data. This technique is widely used in the social sciences but is most noticeable in the field of psychometrics. This method allows grouping variables that are strongly correlated with each other.

The steps to follow in order to carry out the EFA are the following:

- i. Decide the m number that will correspond to the factors to be extracted and which will be the method of extraction. The number of factors can be decided by the Kaiser rule, the 75% of the variance, the elbow rule, the parallel analysis rule, specified by the researcher and by the halving criterion. The extraction method can be the principal components analysis, principal axes factorisation, unweighted least squares method, generalised least squares method, maximum likelihood method, imaging factorisation method and alpha method.

- ii. The next step is to analyse the one-factor solution. If the adjustment of the data is good, it can be concluded that the solution is one-dimensional, and if not, it will be successively tested until there are enough factors for the model.
- iii. The solution obtained is called the initial factorial matrix, which is difficult to interpret and is not the only possible solution.
- iv. This is followed by the rotation method, where the researcher decides whether to perform an orthogonal rotation (Varimax, Quartimax or Equamax) or an oblique rotation (Oblimin or Promax), in order to facilitate interpretation.
- v. Finally, the results are interpreted, indicating the variables that correspond to each factor and the name of each factor.

CONFIRMATORY FACTOR ANALYSIS (CFA)

Confirmatory Factor Analysis tries to delimit whether the number of factors that have been obtained in the EFA and their loadings correspond to what is expected from a previous theory about the data. CFA has three stages:

- i. The first stage specifies the model by establishing that each variable saturates only on the common factor it measures, that the common factors are correlated and the uncorrelated ones uncorrelated.
- ii. The second stage is to identify whether the model is estimable. Each factor must explain at least two items.
- iii. The third stage is to estimate the model. The most classical method is the maximum likelihood method. Once estimated, goodness-of-fit values are observed to see if the model fits the data correctly.

CORRESPONDENCE ANALYSIS (CA)

Correspondence Analysis is a multivariate statistical technique of categorical data analysis used to discover and represent underlying structures in a data set. It allows large amounts of information to be summarised into a small number of dimensions or factors. Correspondence Analysis has two basic objectives:

- Association between categories of columns or rows. It measures the association of only one row or column to see if the modalities of a variable can be combined.
- Association between row and column categories. It studies whether there is a relationship between the categories of variables in the rows and columns.

It is a technique used to represent the rows and columns of a contingency table as points in a low-dimensional vector space, in other words, its main objective is to transform a table with numerical information into a graphical representation that facilitates its interpretation. There are two types:

- Simple Correspondence Analysis (SCA) is the analysis in which the data are in the form of a contingency table and the existing relationships between two variables are evaluated.
- Multiple Correspondence Analysis (MCA) is the analysis in which two or more nominal variables are used.

The MCA attempts to find two matrices called A and B of Cartesian coordinates

$$A = \begin{pmatrix} a'_{1} \\ a'_{2} \\ \dots \\ a'_{k} \end{pmatrix}, \quad B = \begin{pmatrix} b'_{1} \\ b'_{2} \\ \dots \\ b'_{k} \end{pmatrix}$$

where A represents the row points and B the column points. There are different methods to calculate these matrices and they are called standardizations:

- Symmetric. In each dimension the row scores are the weighted average of the column scores between the corresponding eigenvalues. This method treats the rows and columns symmetrically and therefore distributes the inertia identically over the row and column coordinates.
- Principal per row. The Euclidean distances between the row points approximate the chi-square distance between the rows of the table.
- Principal. This method allows to examine the differences or similarities between the categories of variables, but separately.
- Custom. A value included in the interval $[-1, 1]$ is specified, where closer to 1 is equal to the main method per object and closer to -1 is equal to the main method per row/column.

8.5. RESULTS

The questionnaire has an acceptable reliability, with $\alpha = 0.718$. An Exploratory Factor Analysis was carried out with three factors:

The first factor designated "Online Shopping" is composed of the items ONLSUP, ONLINE, PQONLSUP, ONLDES, MESONL, INTONL, COMPONL, PRESONL. The second factor called "Purchase" is composed of the items COMPANTE, NUMSUP, CAMBIO, NUMDES FREQCOM, TIEMPCOMP, PLAN and COMPRES. This last item should belong to the first factor, but with the

data obtained the analysis includes it in this one. The reason may be that, as this item asks about face-to-face shopping, it is related to the factor corresponding to shopping and the supermarket. The third factor, called "Economy", is made up of the items CALIDAD, GASTO, MOTAUM, TIRCOM and GASTDES.

Once the Exploratory Factor Analysis was completed, a Confirmatory Factor Analysis was carried out, obtaining results that indicate that the dimensions proposed by the EFA are acceptable for the data.

The analysis was carried out for the model with three and four factors. In both cases the χ^2 statistic is significant (p -value < 0.05), which is due to its sensitivity to sample size. For this reason, other fit indices are proposed, such as the Goodness of Fit Index (GFI = 0.870), the Adjusted Goodness of Fit Index (AGFI = 0.879), the Root Mean Square Residual Index (SRMR = 0.05), the Adjusted Goodness of Fit Index (AGFI = 0.879) and the Root Mean Square Error (RMR = 0.05), the Root Mean Square Error of Approximation (RMSEA = 0.058), the Comparative Fit Index (CFI = 0.885), the Normalised Fit Index (NFI = 0.790) and the Non-Normalised Fit Index (NNFI or TLI = 0.383).

The indices could be considered unacceptable, but as they are so close to the optimal points, the model is considered to fit the data well.

To create profiles and characterise individuals, the variable "change" is used, which takes into account whether each individual respondent considers that their supermarket habits have changed after Covid-19. The profiles obtained are as follows:

Consumers who believe that their consumption habits in supermarkets have been affected for the worse are individuals between 25 and 45 years of age who, after the start of the pandemic, have seen their income fall by between 10 and 100% and who buy every two weeks, having reduced their frequency for fear of contagion and shopping more slowly to keep spending under control, which has increased due to eating more meals at home and less in restaurants or ordering from home. In terms of the importance they attach to spending, these are individuals who, after the pandemic, are planning to change it because they will start to place more value on prices. With regard to online shopping, they shopped online before the pandemic for convenience and because it was cheaper by avoiding temptation, and for the same reason they shopped more in this way after the pandemic began, even replacing face-to-face shopping with online shopping.

Consumers who consider that their shopping habits in supermarkets have not changed are individuals aged between 46 and 65 who have not seen their income change since the pandemic began and who shop once, twice or three times a week, maintaining the same frequency and regardless of the time spent. The expenditure they expect to make once the pandemic is over is the same as it has been up to now. As regards online shopping, these are consumers who did not shop in supermarkets before the pandemic started and have not changed from face-to-face shopping to online shopping, but have only kept the former.

Consumers who believe that their supermarket spending habits have been affected for the better are individuals aged 18-25 who have seen their income increase since the start of the

pandemic. They have also increased the frequency of shopping, generally shopping every day to look for offers/promotions and different types of brands or products and shopping faster to spend less time in the supermarket, as well as larger, with more products. Considering the expense of shopping, they aim to decrease it through smaller and more organised purchases. These are individuals who started shopping online after the start of the pandemic for fear of contagion but intend to continue shopping both online and in person once the pandemic is over.

8.6. CONCLUSIONS

- The internal consistency of the questionnaire was assessed by Cronbach's Alpha coefficient and this acceptable value was 0.718.
- The validity of the questionnaire was measured with the EFA and launched three factors through the organisation of the different items, and the internal consistency of two of the three dimensions was acceptable.
- The results obtained show that the factor structure have three factors.
- As a result of the CFA it was concluded that the model fitted the data correctly and that the model with three factors was more appropriate than the one with four factors.
- The consumer profiles are obtained by means of different ACS, which return three types of consumers:
 - Those who have experienced a change for the worse, are individuals in general who have lost income and who focus their purchases on money.
 - Those who have not experienced change are individuals who have kept all their habits the same.
 - Those who have experienced a change for the better, are individuals who do not focus so much on money, but more on fear and exposure to the virus.

ESTUDIO DEL COMPORTAMIENTO DEL CONSUMIDOR DURANTE LA COVID-19 EN LOS SUPERMERCADOS

La actual pandemia producida por la COVID-19 nos ha obligado a realizar muchos cambios en nuestra vida por distintos motivos, pero el principal ha sido la crisis económica. Uno de los momentos más recordados fue el principio del estado de alarma, lo que se puede llamar como una fase de “abastecimiento”. Los supermercados comenzaron a abarrotarse de personas desesperadas por llenar sus neveras y despensas. Una vez que pasó esa fase, comenzó la del confinamiento, en la que la rutina de compras de muchos de nosotros se vio afectada.

Esta encuesta se hace con el fin de realizar un estudio sobre como es el nuevo comportamiento de los consumidores en los supermercados tras esta fase de “abastecimiento”. Conocer las nuevas rutinas de los consumidores, saber si se han mantenido leales a su supermercado de confianza, ver si la crisis económica ha afectado en la forma de realizar las compras...

El estudio se efectúa para la realización del Trabajo de Fin de Grado (TFG) del Grado en Estadística. Agradezco mucho tu colaboración.

1. GÉNERO (GENERO)

- Mujer Hombre Prefiero no decirlo

2. EDAD (EDAD)

- 18 – 25 25 – 45 46 – 65 Mayor de 65

3. PROVINCIA DE RESIDENCIA (PROV)

4. ZONA DE RESIDENCIA (ZONA)

- Rural Urbano

5. ¿CON QUIÉN RESIDES DESDE QUE EMPEZÓ LA PANDEMIA? (CONV)

- Yo solo Con mis padres
 Con mi pareja Otros miembros familia

6. ¿CUÁL ES TU SITUACIÓN LABORAL ACTUALMENTE? (LABORAL)

- Estudiante Paro/ ERTE
 Trabajador Desempleado sin prestación
 Jubilado

7. ¿HAN CAMBIADO TUS INGRESOS DESDE ANTES DE LA PANDEMIA? (INGRESOS)

- Si, han disminuido No han cambiado
 Si, han aumentado

8. EN RELACIÓN CON LO QUE GANABAS ANTES DE LA PANDEMIA, ¿CUÁNTO HAN CAMBIADO TUS INGRESOS? (CAMBING)

- Han aumentado mis ingresos
- Han disminuido entre 90-100%
- Han disminuido entre 60-89%
- Han disminuido entre 10-59%
- Han disminuido entre 0-9%
- Se mantienen igual

9. ¿DÓNDE REALIZABAS LAS COMPRAS PRESENCIALES ANTES DE LA PANDEMIA? (COMANT)

- Tienda de barrio/pueblo
- Supermercado

10. ¿REALIZAS LA COMPRA EN UN ÚNICO SUPERMERCADO O EN VARIOS? (NUMSUP)

- En varios, voy buscando los precios más baratos
- En el mismo, quiero reducir el riesgo de contagio
- En el mismo, sólo tengo uno cerca de casa
- Realizo la compra en una o varias tiendas de barrio/ pueblo

11. DESDE QUE COMENZÓ LA PANDEMIA, ¿CUÁNTAS VECES COMPRAS AL MES EN SUPERMERCADOS? (NUMDES)

- Cada dos semanas
- Dos/tres veces por semana
- Una vez a la semana
- Todos los días

12. ¿POR QUÉ MOTIVO HAS CAMBIADO LA FRECUENCIA DE COMPRA? (FREQCOM)

- Ha disminuido porque tengo miedo a contagiarme
- Ha aumentado porque busco ofertas y voy comprando al día
- Ha disminuido solo por pereza
- Compró con la misma frecuencia de antes

13. ¿DEDICAS EL MISMO TIEMPO QUE ANTES A HACER LA COMPRA? (TIEMCOMP)

- No, voy más deprisa para estar menos tiempo en el supermercado
- No, voy más despacio porque pienso mejor la compra para controlar el dinero
- Sí, no me preocupa el tiempo que paso en el supermercado

14. ¿PLANIFICAS LA COMPRA? (PLAN)

- Sí, para pasar menos tiempo en el supermercado
- Sí, para gastar lo menos posible
- No, según voy viendo cosas voy comprando

15. ¿CÓMO ES LA CALIDAD DE LOS PRODUCTOS QUE COMPRAS? (CALIDAD)

- La misma calidad que antes
- Menos calidad, busco que sea más barato
- Más calidad que antes, no me importa el precio

16. ¿HAS VARIADO EL GASTO EN LA COMPRA? (GASTO)

- Ha aumentado más del 40%
- Se mantiene igual
- Ha disminuido más del 40%
- Ha aumentado entre 10-40%
- Ha disminuido entre 10-40 %

17. MOTIVOS POR LOS QUE HA AUMENTADO EL GASTO DE LA COMPRA (MOTAUM)

- Hago compras más grandes (más productos)
- Hago más comidas en casa, no pido comida a domicilio
- Tiro mucha comida, no me organizo bien
- No ha aumentado

18. ¿CUÁNTA COMIDA TIRAS A LA SEMANA DESDE QUE EMPEZÓ LA PANDEMIA? (TIRCOM)

- Tiro mucha, me da igual el gasto
- Tiro poca, lo que no se termina tras comer
- No tiro nada, planifico las compras o hago aprovechamiento

19. ¿CAMBIARÁN TUS GASTOS EN COMPRAS TRAS LA PANDEMIA? (GASTDES)

- Sí, haré compras más pequeñas y organizadas
- Sí, le daré más importancia a los precios
- Sí, volveré a tener más dinero y volveré a gastar más
- Se mantendrán igual

20. ¿UTILIZABAS LA COMPRA ONLINE EN SUPERMERCADOS ANTES DE LA PANDEMIA? (ONLSUP)

- Sí, siempre compraba online
- Sí, pero también hacía compra presencial
- No, nunca compraba online

21. ¿COMPRABAS OTROS PRODUCTOS QUE NO SEAN PRODUCTOS DE ALIMENTACIÓN ANTES DE LA PANDEMIA DE FORMA ONLINE? (ONLINE)

- Compraba, pero no era mi forma habitual
- Todo lo compraba online
- No he comprado nada online

22. ¿POR QUÉ REALIZABAS COMPRAS ONLINE EN SUPERMERCADOS ANTES DE LA PANDEMIA? (PQONLSUP)

- Por comodidad
- Por falta de tiempo
- Me parece más económico porque evito tentaciones
- No compraba online

23. DESDE QUE COMENZÓ LA PANDEMIA, ¿HAS EMPEZADO A COMPRAR ONLINE EN SUPERMERCADOS? (ONLDESP)

- Sí, por miedo a contagiarme
- Sí, me parece más económico porque evito tentaciones
- Sí, dificultades para acceder a un supermercado (falta de coche, confinamientos...)
- No he comprado online desde que empezó la pandemia

24. ¿CUÁNTAS VECES AL MES HACES COMPRA ONLINE EN SUPERMERCADOS DESDE QUE COMENZÓ LA PANDEMIA? (MESONL)

- Cada dos semanas
- Una vez a la semana
- Dos o tres veces por semana
- Todos los días
- No compro online

25. ¿CREES QUE ES INTERESANTE LA COMPRA ONLINE EN SUPERMERCADOS? (INTONL)

- Sí, evito exponerme al virus
- Sí, es más cómodo
- Me gusta comprar online y presencialmente
- No, me parece muy complicado
- No, prefiero ver todos los productos

26. DURANTE EL CONFINAMIENTO Y EN COMPARACIÓN CON LOS MESES ANTERIORES, EN SUPERMERCADOS, ¿HAS COMPRADO PRESENCIALMENTE: (COMPRES)

- Más que antes
- Igual que antes
- Menos que antes

27. DURANTE EL CONFINAMIENTO Y EN COMPARACIÓN CON LOS MESES ANTERIORES, EN SUPERMERCADOS, ¿HAS COMPRADO ONLINE: (COMONL)

- Más que antes
- Igual que antes
- Menos que antes

28. ¿HAS MODIFICADO LA COMPRA PRESENCIAL EN SUPERMERCADOS POR COMPRA ONLINE? (PRESOONL)

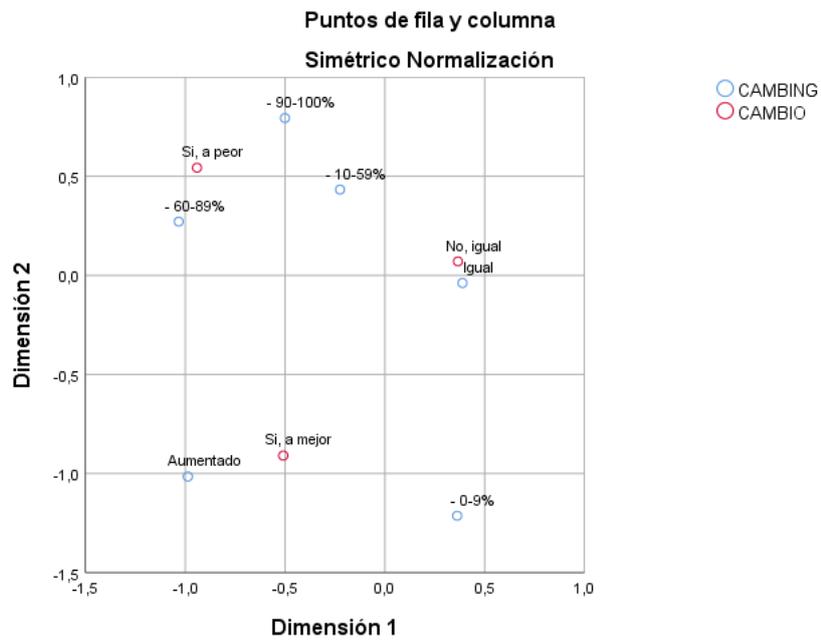
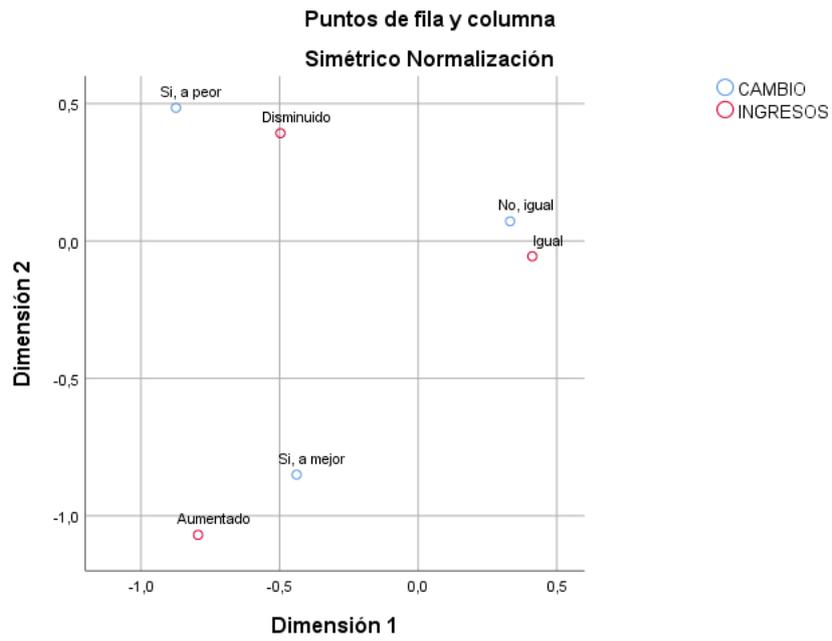
- Sí, solo compro online
- No, solo compro presencialmente

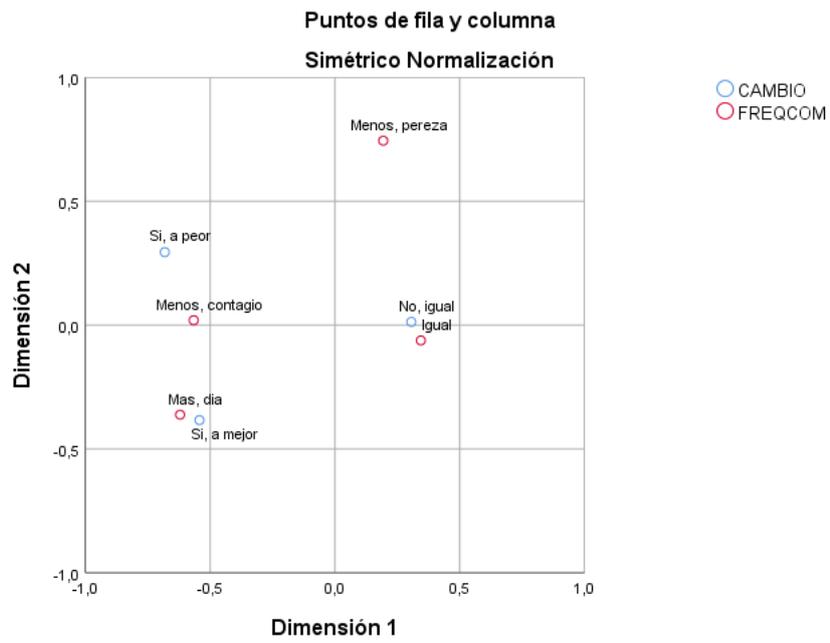
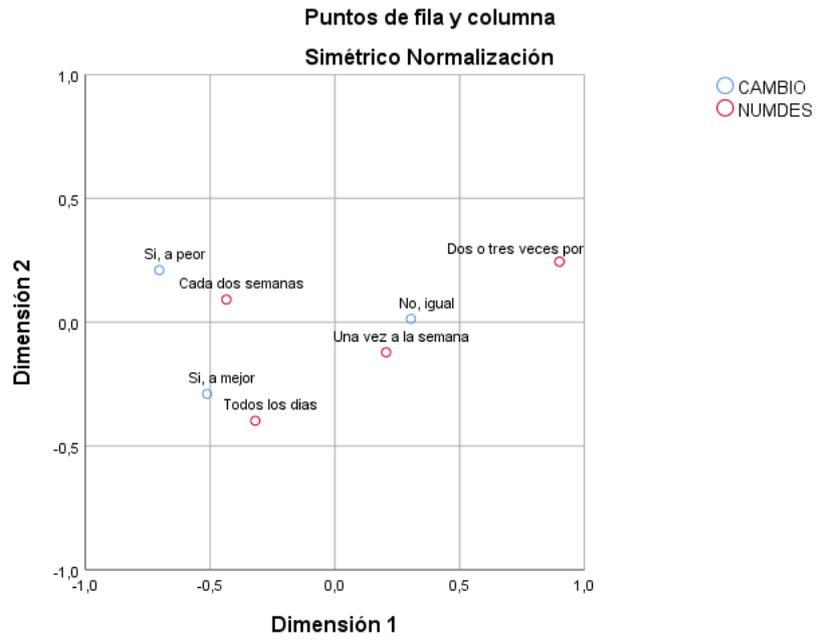
No, compro de las dos formas

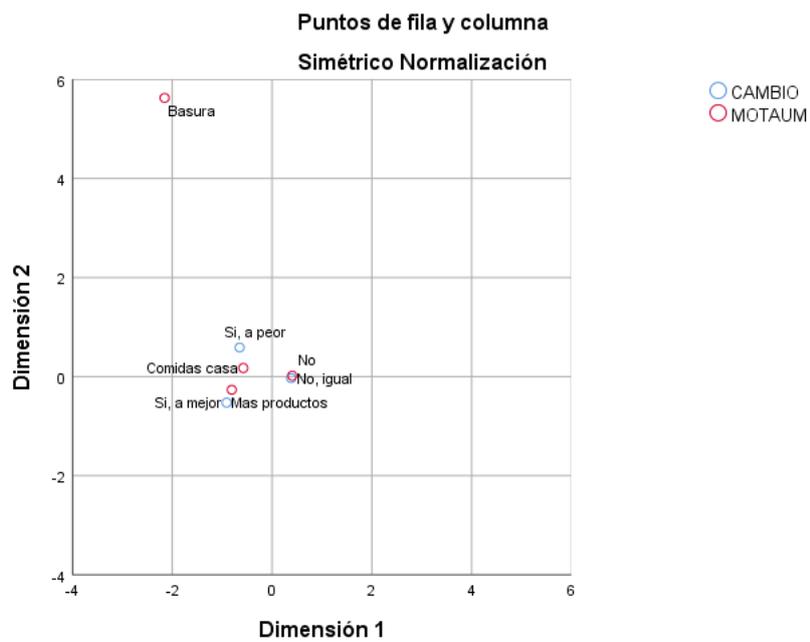
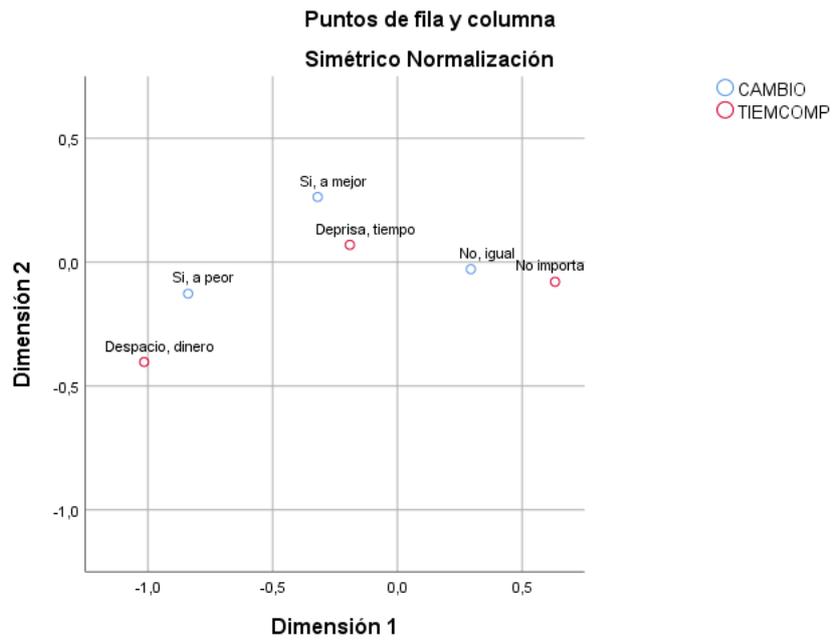
29. ¿CONSIDERAS QUE TUS HÁBITOS DE CONSUMO EN SUPERMERCADOS HA CAMBIADO DESDE QUE COMENZÓ LA PANDEMIA? (CAMBIO)

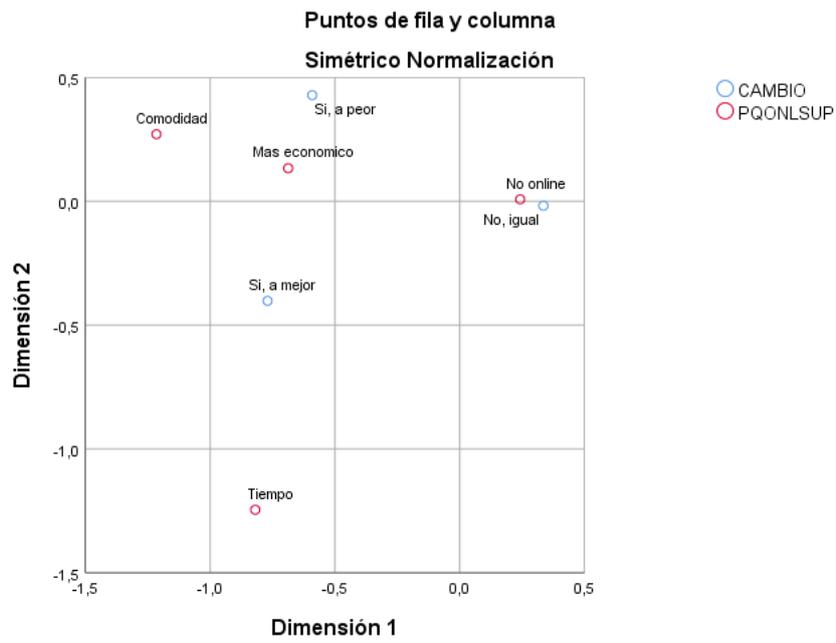
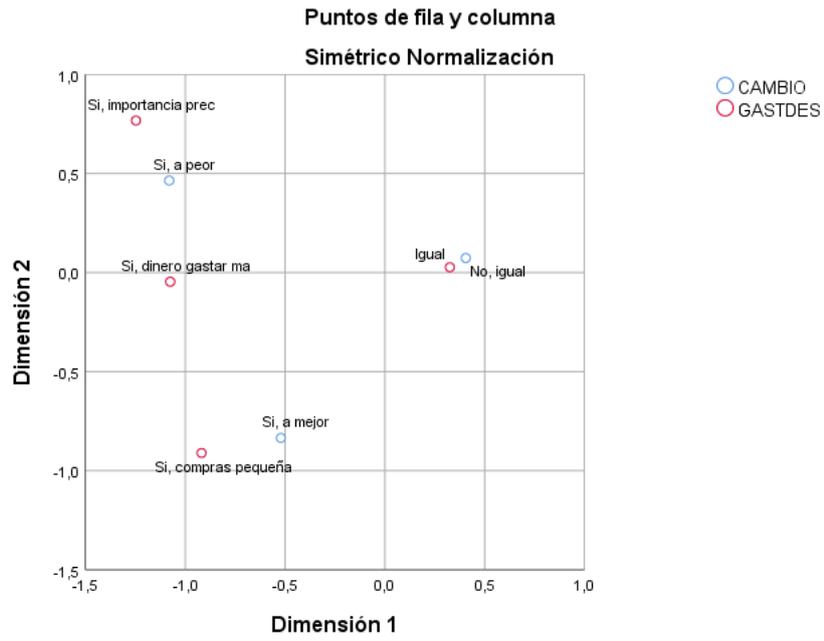
- Sí, se han visto afectados a peor
- Sí, se han visto afectados a mejor
- No, se mantienen igual

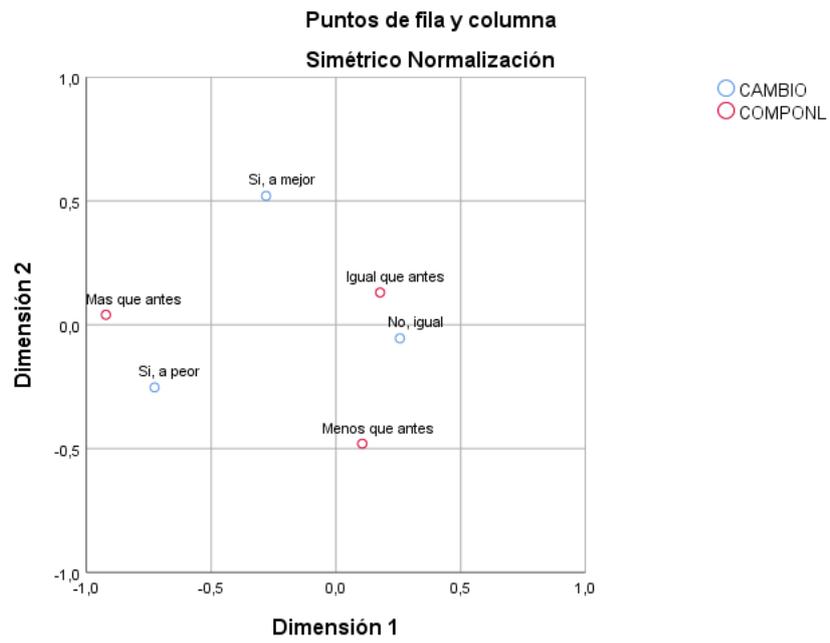
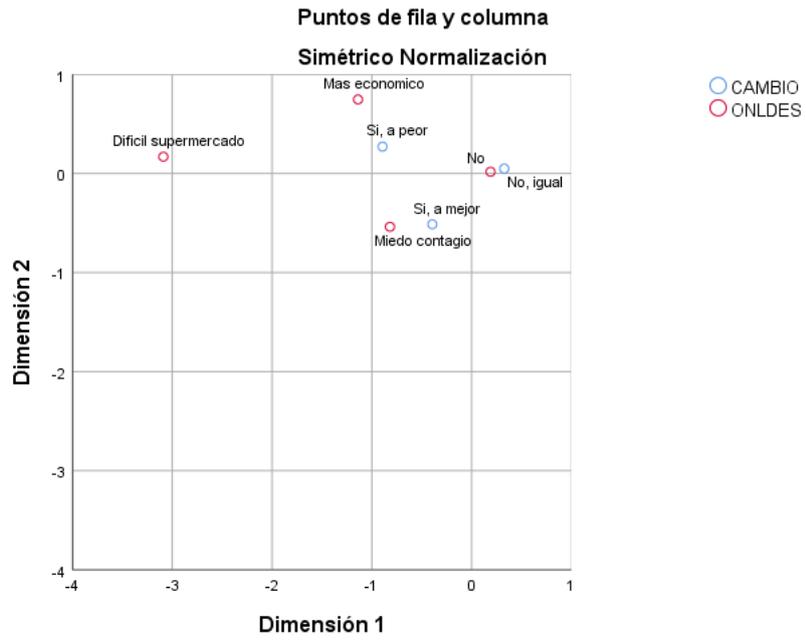
ANEXO II. ANÁLISIS DE CORRESPONDENCIAS, CLARA PALOMAR SEVILLANO











Puntos de fila y columna

Simétrico Normalización

