

UNIVERSITAT DE BARCELONA

Facultat de Ciències Econòmiques I Empresariales

Departament d'Economía i Organització d'Empreses



UNIVERSITAT DE BARCELONA



**INSTRUMENTOS DE MARKETING APLICADOS A LA COMPRA
DE PRODUCTOS ECOLÓGICOS: UN CASO DE ESTUDIO ENTRE
BARCELONA, ESPAÑA Y LA PAZ, MÉXICO**

Doctorando: Lizbeth Salgado Beltrán

Director: Dra. María Esther Subirá Lobera

Co-director: Dr. Luís Felipe Beltrán Morales

Programa de Doctorado: Investigación y
Técnicas de Mercado (bienio 2004-2006).

Barcelona 2009.

PARTE III

ANÁLISIS EMPÍRICO

*Cuando las leyes de la matemática se refieren
a la realidad, no son ciertas;
cuando son ciertas, no se refieren a la realidad*

Albert Einstein (1879-1955)
Científico estadounidense de origen alemán

CAPÍTULO 8

METODOLOGÍA

8.1 Introducción

En la actualidad, los investigadores se ven obligados a recoger un gran número de medidas para poder captar de forma adecuada la complejidad de los fenómenos de las ciencias sociales. Para tratar con este conjunto de muestras se han implantado métodos multivariantes, los cuales permiten analizar simultáneamente conjuntos amplios de variables, como es el caso de las decisiones de compra que anteriormente hemos mencionado, que son efecto de una mezcla de variables más no de una sola o dos. Por lo tanto, la técnica a utilizar en esta investigación, un modelo de ecuaciones estructurales, se considera una extensión de varias técnicas multivariantes como la regresión múltiple, el análisis factorial y análisis de senderos.

En este capítulo se trazan los lineamientos metodológicos en los que recae este estudio. Así mismo, se elabora un cuestionario para obtener la información cuantitativa para comprobar y concretar nuestro modelo de ecuaciones estructurales.

8.2 Metodología de investigación

Con este apartado entramos a la parte empírica propiamente dicha de la presente investigación. Se va a describir en primer lugar el modelo de ecuaciones estructurales, en su forma abreviada SEM por sus siglas en inglés (Structural Equation Modeling), sus características, fases, estrategias de modelización, estimación y evaluación del modelo, así como el diseño del cuestionario y población objeto de estudio.

8.2.2 Modelo de ecuaciones estructurales

El nacimiento de los hoy llamados modelos de ecuaciones estructurales se creó en 1970, año en que el econométra Goldberg organizó una conferencia sobre modelos que analizan la causalidad a la que invitó a estadísticos, psicómetras, econométras y

sociómetras. Se dio por primera vez la misma importancia a la teoría que considera la relación entre indicadores y constructos que a la que se interesa en las relaciones de los constructos entre sí (Batista y Coenders, 2000: 16).

Los modelos de ecuaciones estructurales sirven para el estudio de las relaciones causales entre variables. En nuestra investigación, aportará una de las herramientas más potentes para el estudio de las relaciones causales sobre datos no experimentales cuando estas relaciones son de tipo lineal. Los modelos de ecuaciones estructurales nos permitirán (*Op. Cit.*, 2000:8):

1. Abordar los fenómenos de los instrumentos de marketing ecológico y la decisión de la compra de productos ecológicos en toda su globalidad, teniendo en cuenta su gran complejidad.
2. Simplificar las grandes matrices multivariantes que surjan, que tienen un excesivo volumen de datos.
3. Especificar el modelo propiamente, de acuerdo a nuestro propio criterio y conocimientos, modificándolo de forma flexible según su ajuste a los datos.
4. Eliminar el efecto error de medida de las relaciones entre variables.

El modelo de ecuaciones estructurales, está compuesto por dos sub-modelos: de medida y estructural. El primero, especifica las relaciones que guardan los factores o variables latentes con sus respectivos indicadores, tal y como se especifican las relaciones entre variables observables y latentes en un análisis factorial confirmatorio; es decir, ecuaciones de medida de estos factores, cuya representación en forma de ecuación es:

$$Y = \Lambda_y \eta + \epsilon$$

$$X = \Lambda_x \xi + \delta$$

Donde:

η : vector $m \times 1$ de variables latentes endógenas,

ξ : vector $k \times 1$ de variables latentes exógenas,

Λ_x : matriz $q \times k$ de coeficientes de variables exógenas,

Λ_y : matriz $p \times m$ de coeficientes de variables endógenas,

δ : vector $q \times 1$ de errores de medición para los indicadores exógenos,

ε : vector $p \times 1$ de errores de medición para los indicadores endógenos.

El segundo especifica las relaciones direccionales de las variables latentes entre sí; es decir, son ecuaciones que expresan relaciones entre factores, cuya representación en forma de ecuación es:

$$\eta = B\eta + \Gamma\xi + \zeta$$

Donde:

η : vector $m \times 1$ de variables latentes endógenas,

B : matriz $m \times m$ de coeficientes de las variables endógenas,

Γ : matriz $m \times k$ de coeficientes de las variables exógenas,

ξ : vector $k \times 1$ de variables latentes exógenas,

ζ : vector $m \times 1$ de términos de perturbación aleatoria.

Esta configuración revela que el SEM es la combinación de un modelo psicométrico de análisis factorial y de otro econométrico. Ambos modelos se desarrollan en el siguiente capítulo.

8.2.2.1 Aplicaciones de las características de los modelos de ecuaciones estructurales

El SEM no designa una técnica estadística única sino que se refiere a una familia de procedimientos. Es a priori y requiere que los investigadores piensen en términos de modelos pero no significa que sea exclusivamente confirmatorio. La explícita representación de la distinción entre variables observables y latentes (concepto supuesto y no observado que sólo puede ser aproximado mediante variables medibles u observables) es una característica de muchos modelos de ecuaciones estructurales (Kline, 2005:9).

Otras características son que la estadística básica en el SEM es la covarianza. La covarianza entre dos variables X y Y, es como se muestra:

$$\text{COV}_{xy} = r_{xy} \text{SD}_x \text{SD}_y$$

Donde r_{xy} es la correlación de Pearson entre X y Y y donde, SD_x y SD_y son sus desviaciones estándar. Una covarianza representa la fortaleza de la asociación entre X y Y y sus variabilidades. Algunas veces es referida como una correlación no estandarizada porque no tiene límites sobre sus valores altos y bajos (*Idem*, 2005: 12). Decir que la covarianza es la estadística básica de SEM significa que hay dos objetivos del análisis: entender patrones de correlaciones entre un grupo de variables, y cuanto es posible explicar de su varianza con el modelo especificado por el investigador.

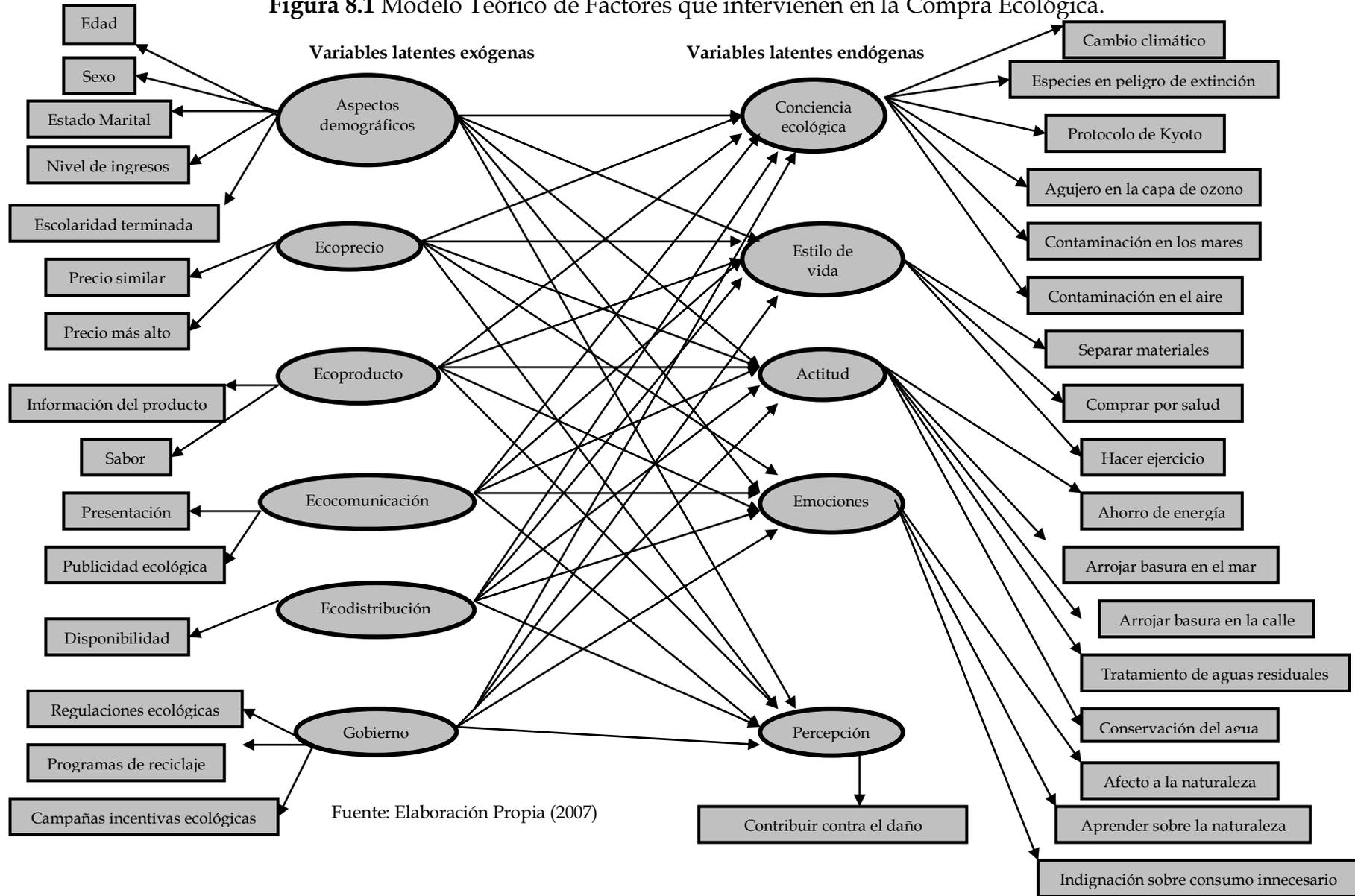
También la técnica del SEM no es solo para datos no experimentales (correlacionales). Y, que la mayoría de las aplicaciones de ecuaciones estructurales se hacen con grandes muestras. Una de las diferencias observables entre los sistemas de ecuaciones estructurales y las otras técnicas de análisis multivariante es la capacidad

de analizar relaciones para cada subconjunto de variables, además de permitir interrelaciones entre algunas variables de estos subconjuntos.

8.2.2.2 Fases de nuestro modelo de ecuaciones estructurales

Con base en la extensa revisión de literatura realizada en el capítulo 2, se ha definido un modelo teórico (figura 8.1) de las variables a utilizar en nuestro modelo final de ecuaciones estructurales sobre la compra ecológica y los instrumentos de marketing, el cual se intentara contrastar estadísticamente y finalizar con una verificación del modelo.

Figura 8.1 Modelo Teórico de Factores que intervienen en la Compra Ecológica.



La falta de confirmación de una o varias hipótesis establecidas en la elaboración del modelo conlleva a la modificación de este, manteniendo un significado lógico-sustantivo, además con significancia estadística. Toda modificación del modelo implica la repetición de todo el análisis estadístico. Para Batista y Coenders (2000: 55) las etapas de construcción de modelos de ecuaciones estructurales son las siguientes:

1. La primera etapa de especificación, traduce en un conjunto de ecuaciones las teorías verbales formuladas previamente y que atañen a:
 - Las variables latentes o dimensiones que deben considerarse.
 - Los efectos entre las variables latentes y su tipo (directo, indirecto, conjunto o espúreo)
 - Los indicadores que asignamos a cada dimensión.
 - Las covarianzas entre variables latentes exógenas.
2. Si se asume que la teoría es correcta se podrían derivar a partir del modelo las varianzas y covarianzas entre las variables observables.
3. Una vez se ha cerciorado de que los parámetros son estimables, pueden recogerse los datos y calcularse las varianzas y covarianzas muestrales.
4. Cuando se dispone de la información muestral y de las relaciones establecidas entre covarianzas y parámetros, puede procederse a la estimación.

5. El modelo especificado a partir de la teoría debe contrastarse con los datos empíricos mediante pruebas de significación adecuadas en la etapa de diagnóstico.
6. Solo una vez verificado con éxito, puede el modelo emplearse en la etapa de utilización para evaluar la intensidad de las relaciones, primero entre factores y sus indicadores y después entre unos factores y otros.

8.2.2.3 Las estrategias de modelización a seguir

Desde el punto de vista teórico, en la Modelización de Ecuaciones Estructurales se estiman coeficientes desconocidos en un conjunto de ecuaciones estructurales lineales. Las variables que intervienen en el sistema de ecuaciones son: variables observadas directamente y variables llamadas latentes o constructos, que sólo pueden ser medidas a través de otras variables directamente observables. La Modelización de Ecuaciones Estructurales supone que existe una estructura causal entre un conjunto de variables latentes, y que las variables observadas son indicadores de las variables latentes. Las variables latentes pueden aparecer como combinaciones lineales de las variables observadas.

La flexibilidad de los sistemas de ecuaciones estructurales nos permitirá utilizar la técnica de múltiple formas y en función de los objetivos a seguir. Empero, debemos seleccionar al inicio del estudio qué estrategia queremos seguir para alcanzar nuestro objetivo; es decir, un plan de acción orientado a la consecución de resultados. Jöreskog (1993) distingue tres estrategias que se pueden adoptar en relación con los modelos de ecuaciones estructurales:

1. de modelización confirmatoria: es la aplicación más directa de los SEM, donde el investigador especifica un modelo aislado y dicho modelo se utiliza para evaluar su significancia estadística.
2. de modelos rivales: permite evaluar el modelo estimado con modelos alternativos; esta estrategia sirve para realizar comparaciones y escoger el más idóneo y,
3. desarrollo del modelo: el investigador que sigue esta estrategia trata de re-especificar un modelo básico de teoría para encontrar uno nuevo. El modelo alterado es entonces probado de nuevo con los mismos datos. El objetivo de este proceso es descubrir un modelo con dos propiedades: hacerlo en el sentido teórico y su correspondencia estadística de los datos es razonable.

En el presente análisis se utilizó la estrategia de desarrollo del modelo, porque es lo que más se adecua a nuestros objetivos.

8.2.2.4 La representación gráfica del modelo de ecuaciones estructurales

El SEM representa las relaciones estructurales mostrando las líneas de causalidad entre las variables latentes endógenas. La técnica se sirve de grafos que reflejan el proceso causal atendiendo a ciertas convenciones que los hacen acordes con las ecuaciones (Batista y Coenders, 2000:23)

1. La relación causal entre variables se indica por una flecha \rightarrow cuyo sentido es desde la variable causa hacia la variable efecto. Cuando entre dos variables no hay flecha alguna significa que su efecto es cero. Las variables a las que llega alguna flecha de denominan *endógenas* y aquellas a las que no llega ninguna flecha, *exógenas*.

2. La covarianza entre dos variables exógenas o dos términos de perturbación sin interpretación causal alguna se representa por una flecha bidireccional \longleftrightarrow que las une en ambos sentidos.

3. Cuando una teoría incluye tanto variables observables como latentes, las primeras se enmarcan en cuadros \square y las segundas en círculos \bigcirc . Las llamadas relaciones de medición conectan las variables observables con las latentes, con flechas unidireccionales con origen en las latentes.

Los componentes del SEM son:

Variables latentes: endógenas η_1, η_2, η_3 , exógenas ξ_1, ξ_2 ,

Variables observadas: endógenas $Y_1, Y_2, Y_3, Y_4, Y_5, Y_6$, exógenas $X_1, X_2, X_3, X_4, X_5, X_6$,

Errores de medida: de variables observadas endógenas $\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5, \varepsilon_6$, de variables observadas exógenas $\delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \delta_6$,

Coefficientes de correlación: $\theta_\varepsilon, \theta_\delta$, correlacionan a los errores de medida.

Términos de perturbación: $\zeta_1, \zeta_2, \zeta_3$, los cuales incluyen los efectos de las variables omitidas, los errores de medida y la aleatoriedad del proceso especificado. La variación en el término de perturbación se denota por ψ y la covariación entre los términos de perturbación i-ésimo y j-ésimo se denota por ψ_{ij} .

Coefficientes de regresión: λ_x, λ_y , que relacionan las variables latentes con las observadas.

Coefficientes de regresión: γ_{11} , γ_{22} (representa la relación entre una variable latente exógena y una endógena), β_{21} , β_{22} , β_{31} (relacionan las variables latentes endógenas entre sí), y ϕ_{21} (representa la covariación entre las variables latentes exógenas).

8.2.2.5 La identificación del modelo de ecuaciones estructurales

Un modelo se dice que está identificado, si los parámetros del modelo completo (modelo estructural y modelo de medida, juntos) pueden estimarse a partir de los elementos de la matriz de covarianza de las variables observadas. La regla del conteo se utiliza para identificar los modelos de ecuaciones estructurales. Se denotará al número total de variables con $s = p + q$, siendo p las variables endógenas y q las exógenas, luego el número de elementos no redundantes en Σ es igual a $\frac{1}{2} s(s+1)$. Además, se denota al número total de parámetros a ser estimados en el modelo como t ; entonces, para realizar la identificación del modelo se debe tener la siguiente condición necesaria $t \leq \frac{1}{2} s(s+1)$.

- Si se tiene la igualdad, se dice que el modelo está *identificado*.
- Si t es estrictamente menor que $\frac{1}{2} s(s+1)$, se dice que el modelo está *sobre identificado*.
- Si t es mayor que $\frac{1}{2} s(s+1)$, entonces el modelo *no está identificado*.

Cuando el modelo de ecuaciones estructurales es recursivo, está también identificado.

8.2.2.6 La estimación del modelo

Partiendo de que el modelo está identificado, el siguiente paso es realizar la estimación del modelo de ecuaciones estructurales. Los modelos de ecuaciones

estructurales pueden usar como datos iniciales la matriz de varianzas-covarianzas o la matriz de correlaciones de las variables observadas. Además de las ecuaciones de los submodelos estructural y de medida que conforman un modelo de ecuaciones estructurales, se debe considerar al vector de parámetros, denotado como:

$$\Omega: \Omega = (\Lambda_y, \Lambda_x, \Theta_\varepsilon, \Theta_\delta, \Phi, B, \Gamma, \Psi)$$

El objetivo es obtener las estimaciones del vector de parámetros Ω , denotado como $\hat{\Omega}$, que minimice la función de discrepancia $F(S, \hat{\Sigma})$ donde $\hat{\Sigma} = \Sigma(\hat{\Omega})$ es la matriz de varianza-covarianza de las estimaciones del modelo llamada matriz de varianza-covarianza ajustada.

La función $F(S, \hat{\Sigma})$ es un escalar que mide la discrepancia (distancia) entre la matriz de varianza-covarianza muestral S y la matriz de varianza-covarianza ajustada $\hat{\Sigma}$.

Esta función de ajuste vendrá dada por la siguiente expresión:

$$F = [S - \Sigma(\Omega)]' W^{-1} [S - \Sigma(\Omega)]$$

Donde:

S : es la matriz de varianzas-covarianzas de la muestra,

$\Sigma(\Omega)$: es la matriz de varianzas-covarianzas predichas por el modelo,

W : es una matriz de ponderaciones que puede tomar diversas formas dependiendo de la distribución que tengan las variables observadas.

Si se asume que la distribución muestral de dichas variables es normal multivariante, entonces la función de ajuste tomará la siguiente forma:

$$F_{\text{NORMAL}} = 2^{-1} \text{Traza}[(S - \Sigma(\Omega))W_2]^2$$

Donde:

W_2 : es una matriz que puede tomar diversas formas en función del tipo de método de estimación que se escoja:

$W_2 = \Sigma^{-1}$: Máxima verosimilitud,

$W_2 = S^{-1}$: Mínimos cuadrados generalizados,

$W_2 = I$: Mínimos cuadrados no ponderados.

La función de discrepancia para los métodos más usuales son:

Máxima verosimilitud:

$$F_{\text{MV}} = \log |\Sigma(\Omega)| + tr[S\Sigma^{-1}(\Omega)] - \log |S| - t$$

Mínimos cuadrados generalizados:

$$F_{\text{MCG}} = 1/2 \text{tr}(I - S^{-1}\Sigma(\Omega))^2$$

Mínimos cuadrados no ponderados:

$$F_{\text{MCNP}} = 1/2 \text{tr}(S - \Sigma(\Omega))^2$$

Además de la técnica de estimación empleada, es necesario elegir entre varios procesos de estimación, como la estimación directa, el bootstrapping, la simulación y el análisis de jackknife. El primero es el más utilizado, en el cual se estima directamente un modelo con un procedimiento de estimación elegido. El segundo,

Bootstrap es una técnica estadística computacional propuesta originalmente por B. Efron a finales de 1970. Puede ser utilizada para estimar la distribución muestral del estimador o del test estadístico, y es particularmente útil donde no se conoce o no está disponible la distribución teórica.

El análisis procede ya sea por medio de un re-muestreo de los datos originales o del modelo estimado con los datos. Tiene una amplia aplicación cuando la teoría asintótica resulta limitada como guía al verdadero, pero desconocido, procedimiento de inferencia muestral finito, en particular cuando desconocemos cómo es de buena la aproximación en la situación práctica planteada (Kline, 2005: 42). El método básico de bootstrap es aplicable cuando se asume que los datos son de muestras independientes, o por lo menos intercambiables, de alguna población. Cuando los datos no pueden ser tratados como independientes (por ejemplo, datos de series de tiempo dependientes), están disponibles métodos más sofisticados.

El tercero, para que el desarrollo de un sistema de simulación sea rentable, debe ser lo suficientemente flexible como para poder incorporar aquellas variaciones que se produzcan por la necesaria modernización o por la introducción de elementos nuevos. Por ello, para hacer frente a tales vicisitudes, los sistemas de simulación deben implementar un mecanismo de validación asignándoles puntuaciones según criterios preferiblemente objetivos, de forma que no se desvirtúen los resultados de los ejercicios simulados. Y por último, el análisis Jacknife, también es conocido como “Quenouille-Tukey Jacknife”. Mosteller & Tukey (1977) la describen así:

“Su nombre intenta sugerir la gran utilidad de una técnica como sustituto de otras herramientas especializadas no disponibles, tal como la fiel herramienta de los Boy Scout, la cual tiene diversos usos... La idea básica de esta prueba consiste en calcular el efecto de cada grupo en que los datos han sido divididos, no tomando el resultado de ese grupo individualmente... sino tomando el efecto del conjunto de datos que resulta de omitir dicho grupo.”

En otras palabras, dada una población de valores obtenidas de los datos observados, se muestrea aleatoriamente dicha población, sin reasignar, hasta llenar los diferentes grupos con un número menor de valores que el grupo original. Es decir, sólo un subgrupo de valores originales son reasignados aleatoriamente a través del muestreo sin reasignación; el resto son eliminados. Estos últimos tres procesos no descansan en un único modelo de estimación del parámetro calculado y sus intervalos de confianza basados en estimaciones múltiples.

8.2.2.7 La evaluación del modelo

Una vez que el modelo ha sido identificado y estimado, el siguiente paso consistirá en evaluar lo bien que nuestros datos se han ajustado al modelo propuesto. Existen tres alternativas para evaluar el modelo: evaluación del ajuste del modelo global, evaluación del ajuste del modelo de medida y evaluación del ajuste del modelo estructural.

1. **Ajuste del modelo global.** Existirá un ajuste perfecto cuando haya una correspondencia perfecta entre la matriz reproducida por el modelo y la matriz de observaciones. Existen tres tipos de medidas de ajuste global: medidas absolutas de ajuste, medidas incrementales de ajuste, y medidas de ajuste de parsimonia.

- Las medidas absolutas de ajuste determinan el grado en que el modelo globalmente (modelo de medida y modelo estructural) predice la matriz de datos inicial. Las principales medidas absolutas de ajuste empleadas son las siguientes:

Estadístico ji-cuadrado, comprueba la significación de la prueba: $H_0: S = \Sigma$, $H_1: S \neq \Sigma$. Para no rechazar la hipótesis nula, el nivel de significación debe ser superior a 0,05.

Estadístico ji-cuadrado no centrado (NCP), es igual al estadístico ji-cuadrado corregido por los grados de libertad. Se consideran aceptables valores lo más próximos posibles a 0.

Raíz cuadrada del error medio cuadrático (RMSEA), (Steiger, 1990); aquí, la discrepancia entre ambas matrices está medida en términos de la población y no en términos de la muestra. Valores inferiores a 0,08 son indicativos de un buen ajuste del modelo en la población.

Índice de bondad de ajuste (GFI), (Jöreskog y Sörbom, 1986) es un índice de la variabilidad explicada por el modelo, oscilando sus valores entre 0 (pobre ajuste) y 1 (perfecto ajuste). Valores superiores a 0,90 indican un ajuste aceptable.

Índice de bondad de ajuste relativo (RGFI), sirve para evaluar la bondad del ajuste del modelo, teniendo en cuenta el tamaño muestral y el número de indicadores. En la práctica, se consideran adecuados los modelos que tienen un RGFI superior a 0,90.

- Las medidas incrementales de ajuste comparan el modelo propuesto con un modelo nulo o básico que se toma de referencia y que, tradicionalmente, suele ser aquel que estipula una falta absoluta de asociación entre las variables del modelo; se trata, por lo tanto, de comparar nuestro modelo con el peor modelo posible. Dentro de estos índices incrementales se puede destacar los siguientes:

Índice de Bondad de Ajuste Ajustado (AGFI), (Jöreskog y Sörbom, 1986) es el GFI ajustado por los grados de libertad del modelo propuesto y del modelo nulo. En la experiencia práctica se considera que valores superiores a 0,90 son indicativos de un buen ajuste del modelo a los datos.

Índice de Bondad de Ajuste Ajustado Relativo (RAGFI), evalúa la bondad de ajuste del modelo. Para la aceptación del modelo se recomienda valores superiores a 0,80.

Índice de Ajuste Normado (NFI), (Bentler y Bonett, 1980) mide la reducción proporcional en la función de ajuste cuando se pasa del modelo nulo al propuesto. El rango de variación de este índice está entre 0 y 1, siendo recomendables valores superiores a 0,90.

Índice de Ajuste No Normado (NNFI) ó Índice de Tucker-Lewis (TLI), (Tucker y Lewis, 1973) compara el ajuste por grado de libertad del modelo propuesto y nulo. Este índice tiende a 1 para modelos con muy buen ajuste, considerándose aceptables valores superiores a 0,90.

Índice de Ajuste Incremental (IFI), (Bollen, 1989) si se tiene dos modelos con los mismos valores para la chi-cuadrado del modelo nulo y propuesto, el que tenga menos parámetros presentará un valor más alto de IFI, siendo el más adecuado. Se consideran aceptables valores próximos a 1.

Índice de Ajuste Relativo (RFI), (Bollen, 1986) este índice proporciona valores próximos a 1 a medida que el modelo va alcanzando un buen ajuste.

Índice de Ajuste Comparativo (CFI), (Bentler, 1990) indica un buen ajuste del modelo para valores próximos a 1.

- Las medidas de ajuste de parsimonia; la parsimonia de un modelo es el grado en que alcanza ajuste para cada coeficiente o parámetro estimado. No se dispone de ninguna prueba estadística asociada de estos índices, por lo que su uso es más adecuado comparando modelos alternativos. Dentro de estas medidas se pueden destacar las siguientes:

Índice de Ajuste Normado Parsimonioso (PNFI), (James *et al*, 1982) es similar al NFI, pero teniendo en cuenta el número de grados de libertad usados para alcanzar el nivel de ajuste. Cuando se comparan modelos alternativos, diferencias en sus valores de este índice entre 0,06 y 0,09 resultan importantes.

Índice de Bondad de Ajuste Parsimonioso (PGFI), (Mulaik, 1989) consiste en el ajuste del GFI basado en la parsimonia del modelo estimado. Son preferibles valores altos de este índice.

Criterio de Información de Akaike (AIC) (Akaike, 1974) sirve para comparar modelos que poseen diferente número de variables latentes. Valores pequeños de esta medida indican una alta parsimonia. (CAIC) (Bozdogan, 1987:360) es una transformación del AIC teniendo las mismas implicaciones.

N Crítico (CN), (Hoetler, 1983) sugiere el tamaño que una muestra debe alcanzar en orden a aceptar el ajuste de un modelo dado sobre una base estadística. Se recomienda valores de al menos 200 para este índice.

2. Ajuste del modelo de medida. Para realizar el ajuste del modelo de medida, el paso inicial consiste en examinar la significación estadística de cada carga obtenida entre el indicador y la variable latente. Una vez comprobada la significación de las cargas, el siguiente paso es comprobar la fiabilidad de cada uno de los indicadores, así como la fiabilidad compuesta del constructo. La fiabilidad para cada indicador será la proporción de varianza que tiene en común con el constructo. Se considera que un indicador debería tener al menos un 50% de su varianza en común con la variable latente. Por lo tanto, el límite que se considera aceptable para esta medida de la fiabilidad compuesta es de 0,50 (Sharma, 1996). Otra medida que normalmente se utiliza para evaluar el ajuste del modelo de medida es la

varianza extraída. Indica la cantidad global de varianza en los indicadores explicada por la variable latente. Si este valor es alto (superior a 0,50), se considera que los indicadores miden adecuadamente dicha variable latente.

3. Ajuste del modelo estructural. Lo primero a analizar en un modelo estructural es la significación alcanzada por los coeficientes estimados. Así, cualquier parámetro estimado debe ser estadísticamente diferente de cero, o lo que es igual, si consideramos un nivel de significación de 0,05, el valor t ha de alcanzar 1,96. Un parámetro no significativo indicaría que la relación propuesta no tiene ningún efecto sustancial, por lo que debería ser eliminado y el modelo reformulado. La interpretación del modelo se hará con arreglo a la estructura teórica en que se ha basado su especificación y a los diversos coeficientes o parámetros estimados, analizando si se corresponden en magnitud y en sentido (positivo o negativo) con las propuestas planteadas por la teoría. La magnitud de los coeficientes no está únicamente determinada por su significación estadística, puesto que depende además de otros factores como el tamaño muestral y la varianza de las variables dependientes e independientes (cuanto mayor es la magnitud de la relación y el tamaño muestral y cuanto menor es la varianza de las variables dependientes e independientes, mayor es la probabilidad de obtener una relación estadísticamente significativa).

Frecuentemente, el proceso de evaluación del modelo desemboca en la modificación del mismo, para lo cual el programa computacional AMOS 7.0 (Análisis of Moment Structure) proporciona ayuda a través de una serie de indicadores. Es importante señalar que nunca se debe hacer modificaciones de un modelo sin que se tenga una explicación basada en la teoría.

8.3 Justificación de la aplicación del modelo de ecuaciones estructurales

Se justifica la aplicación del modelo de ecuaciones estructurales en esta investigación, ya que permite testar modelos teóricos con datos empíricos. Existen procedimientos estadísticos que no ofrecen una forma conveniente para diferenciar entre variables observables y latentes. Como es el Análisis de la Varianza conocido como ANOVA. Este análisis sirve para comparar si los valores de un conjunto de datos numéricos son significativamente distintos a los valores de otro o más conjuntos de datos. El procedimiento para comparar estos valores está basado en la varianza global observada en los grupos de datos numéricos a comparar. Típicamente, el análisis de varianza se utiliza para asociar una probabilidad a la conclusión de que la media de un grupo de puntuaciones es distinta de la media de otro grupo de puntuaciones.

Otro análisis es el factorial, que es una técnica que consiste en resumir la información contenida en una matriz de datos con V variables. Para ello, se identifican un reducido número de factores F , siendo el número de factores menor que el número de variables. Los factores representan a las variables originales, con una pérdida mínima de información. Este modelo matemático es parecido al de la regresión múltiple. Cada variable se expresa como una combinación lineal de factores no directamente observables.

Y la regresión múltiple es una técnica de análisis multivariable en la que se establece una relación funcional entre una variable dependiente o a explicar y una serie de variables independientes o explicativas, en la que se estiman los coeficientes de regresión que determinan el efecto que las variaciones de las variables independientes tienen sobre el comportamiento de la variable dependiente.

Estos modelos son afectados por significancias e intercorrelaciones entre variables observables, pero también ofrecen procedimientos de forma sencilla para probar hipótesis a un alto nivel de extracción.

Sin embargo, la aplicación del modelo de ecuaciones estructurales es considerada una extensión de varias técnicas multivariantes como las antes descritas, que puede construir rápidamente modelos de actitudes y comportamientos que reflejen realmente las relaciones complejas. Es decir, permite dar apoyo al estudio, ampliando métodos de análisis estándar multivariante, incluyendo la regresión, análisis de factores, correlación y el análisis de la varianza. Por lo tanto, el modelo de ecuaciones estructurales resulta conveniente para este estudio; algunas distinciones entre las variables observables y latentes son las siguientes (Kline, 2005:12):

1. Los modelos no necesariamente hipotetizan las variables latentes. La evaluación de modelos que afectan solamente entre variables observables es posible en SEM.
2. Hay más de un tipo de variable latente.
3. Las variables latentes en SEM pueden representar un amplio rango de fenómeno.
4. La distinción latente-observable también provee una forma para tener en cuenta la fiabilidad de la puntuación imperfecta.

8.4 Diseño del cuestionario

Las variables socio-demográficas son frecuentemente la mejor manera para comenzar estudios de segmentación debido a la mucha información publicada disponible y fácil de obtener (Diamantopoulos *et al*, 2003: 466). Las variables sociodemográficas a estudiar fueron edad, sexo, estado civil, escolaridad terminada y nivel de ingresos mensuales. El tipo de encuesta fue estructurado (ver anexo 1 y 2.). La medida utilizada fueron los cinco puntos de la escala Likert. Esta técnica utiliza una categorización del continuo actitudinal graduada según la intensidad (Sarabia, 1999: 99).

Al utilizar escalas se utiliza el método axiomático, *Aplicar el método axiomático a un ámbito de la realidad supone organizar nuestro haber acerca de ese ámbito en forma de teoría axiomática* (Mosterín, 1984: 175); ésta ha estado asociada tradicionalmente con

el escalamiento de estímulos. Para medir el constructo de la compra ecológica se elaboraron 33 ítems. Para ser más específicos, las escalas utilizadas son aditivas, siendo éstas constituidas por una serie de ítems ante los cuales se solicita la reacción del sujeto. El interrogado señala su grado de desacuerdo o acuerdo con cada ítem, respondiéndose con base en la escala:

1=total desacuerdo

2= algo en desacuerdo

3=indiferente

4= algo de acuerdo

5=total acuerdo.

Así, las variables latentes exógenas (independientes) están compuestas por:

1. Aspectos demográficos y sus variables observables son:

- Edad: algunos estudios (Buttel, 1979; Honnold, 1984; Lyons y Breakwell, 1994) han relacionado esta variable con la compra ecológica. Las edades consideradas en el estudio fueron de 18 a 65 años. Es importante mencionar que en las ciudades elegidas existen diferencias, ya que en México su población es más joven que en España.
- Sexo: nos puede arrojar datos interesantes sobre qué sexo tiende más a tomar la decisión sobre la compra ecológica.
- Estado civil: no se han encontrado estudios que relacionen esta variable con la toma de decisión de la compra ecológica; sin embargo, se ha considerado para este estudio.
- Escolaridad terminada: la educación es una parte significativa en cuestión de aspectos ambientales; por lo tanto, nos puede aportar información relevante a nuestro análisis.

- Nivel de ingresos mensuales: acorde con las diferencias económicas que existen entre México, como país en desarrollo y España, como desarrollado, se adaptaron los salarios de ambas ciudades, convirtiéndolos en euros, como se muestra en el siguiente cuadro:

Cuadro 8.1 Nivel de Ingresos Mensuales España-México

Nivel de Ingresos Mensuales:						
España	Menos de 500€	500-1,500€	1,501-2,500€	2,501-3,500€	3,501-5,000€	Más de 5,000€
México	Menos de 129€ (\$2,000 MXN)	129€-256€ (\$2,000-4,000 MXN)	257€- 512€ (\$4,001-8,000 MXN)	513€ -896€ (\$8,001-14,000 MXN)	897€- 1280€ (\$14,001-20,000 MXN)	Más de 1280€ (\$20,000 MXN)

Fuente: Elaboración propia (2007). Tipo de cambio al 13 de diciembre 1 EUR = 15.6143 MXN.

2. Ecoprecio y sus variables observables son:

- Precio similar: esta variable se refiere a si los consumidores están dispuestos a pagar un precio similar de un producto ecológico a uno convencional.
- Precio más alto: da información sobre cuántas personas estarían dispuestas a pagar un precio más elevado por un producto ecológico sobre uno convencional.

3. Ecoproducto y sus variables observables son:

- Información del producto: indica si las personas buscan información sobre los efectos del ambiente de los productos antes de la compra.
- Sabor: una de las variables más importantes en relación con los productos ecológicos pues esto los difiere en gran medida de los convencionales.

4. Ecocomunicación y sus variables observables son:

- Presentación: aspecto importante a conocer, ya que los productos ecológicos normalmente tienen una apariencia menos estética que los convencionales como es el caso de las hortalizas, productos que se venden más en México de cultivo orgánico.
- Publicidad ecológica: con esta variable podemos saber qué tan substancial es para la toma de decisión de la compra ecológica, sobre todo considerando la credibilidad que le dan los consumidores.

5. Ecodistribución y sus variables observables son:

- Disponibilidad: la disponibilidad puede ser un factor determinante para que compre el producto ecológico; no es muy fácil encontrarlo principalmente en la Ciudad de La Paz, donde no existen tiendas especializadas sobre este tipo de producto.

6. Gobierno y sus variables observables son:

- Regulaciones ecológicas: por ejemplo, las normas o ecoetiquetados pueden influir en la decisión de compra de los productos al generar más confianza en los consumidores, dando más información sobre lo que están comprando.
- Programas de reciclaje: esta variable puede servir para conocer la opinión que tienen las personas al respecto. Cabe mencionar que existe una diferencia en las ciudades participantes; en Barcelona existen programas de reciclaje por parte del Gobierno; sin embargo, en La Paz no lo hay; existe de manera voluntaria en empresas que se dedican al reciclaje de aluminio, plástico y cartón principalmente.

- Campanas incentivas ecológicas: puede servir como referencia sobre el papel que está haciendo el gobierno en informar sobre el cuidado del ambiente.

En las variables latentes endógenas (dependientes) nos encontramos con las siguientes:

7. Conciencia ecológica y sus variables observables son:

- Cambio climático: cada uno de los siguientes factores, nos sirve para relacionar el conocimiento de algunos fenómenos ambientales o afines con la toma de decisión de la compra ecológica.
- Especies en peligro de extinción,
- Protocolo de Kyoto,
- Agujero en la capa de ozono,
- Contaminación en los mares,
- Contaminación en el aire.

8. Estilo de vida y sus variables observables son:

- Separar los materiales: esta variable está relacionada con los programas de reciclaje, pero desde el punto de vista de las personas como hábito.

- Comprar por salud: si el consumidor está informado sobre los beneficios de consumir un producto ecológico (sin pesticidas) sabrá que le ayudará en su salud.
- Hacer ejercicio: otro aspecto de la vida saludable que puede ser una causa para que se efectúe la compra de un producto ecológico.

9. Actitudes y sus variables observables son:

- Ahorro de energía: los siguiente ítems evalúan algunas actitudes a favor del ambiente.
- Arrojar basura en los mares,
- Arrojar basura en la calles,
- Tratamiento de aguas residuales,
- Conservación del agua.

10. Emociones y sus variables observables son:

- Afecto a la naturaleza: el tener sentimientos positivos sobre la naturaleza puede desencadenar motivos para realizar la compra ecológica, protegiendo de esta manera la naturaleza.
- Aprender sobre la naturaleza: así mismo, ésta variable puede indicar si el aprendizaje produce que se tome la decisión de compra ecológica.

- Indignación por el consumo innecesario: puede mostrar si las personas sienten respeto por los recursos naturales al indignarse cuando se hace esto de manera superflua.

11. Percepciones y sus variables observables son:

- Cree que al comprar un producto ecológico contribuirá con el daño ecológico.

8.5 Selección de la muestra y error de muestreo

El muestreo es importante, porque a través de él podemos hacer análisis de situaciones de una empresa o de algún campo de la sociedad. El tamaño de la muestra es muy importante; para propósitos descriptivos con menos de 100 casos, debería considerarse pequeña, entre 100 y 200 mediana, más de 200 se consideran grandes (Kline, 2005:15). El muestreo se realizó de forma intencional por cuotas en los establecimientos comerciales más importantes de la Ciudad de La Paz BCS, México y Barcelona, España, donde los participantes contestaron voluntariamente la encuesta.

La aplicación de este tipo de encuesta denominada encuesta personal o por clientela está orientada a localizar al entrevistado, al mismo tiempo que se pretende que sea el cliente real o potencial. Las estimaciones de una encuesta pueden ser afectadas por dos tipos de errores: (1) errores no muestrales que consisten en errores hechos en la recolección y procesamiento de datos y (2) errores de muestreo, que resultan del hecho de haberse entrevistado solo una muestra y no la población completa.

Se utilizó la fórmula para poblaciones infinitas, considerando que ambas ciudades tienen población de más de 100.000 habitantes, segmentando la población de 18 a 65 años como se muestra a continuación.

Fórmula para poblaciones infinitas (más de 100.000 habitantes):

$$n = \frac{Z^2 \times P \times Q}{E^2}$$

Donde:

- n** Tamaño muestral
- Z** Valor correspondiente a la distribución de Gauss 1,96 para $\alpha = 0,05$ y 2,58 para $\alpha = 0,01$.
- p** Prevalencia esperada del parámetro a evaluar. En caso de desconocerse, aplicar la opción más desfavorable ($p=0,5$), que hace mayor el tamaño muestral.
- q** $1-p$ (Si $p=30\%$, $q=70\%$)
- E** Error que se prevé cometer. Por ejemplo, para un error del 10%, introduciremos en la fórmula el valor 0,1. Así, con un error del 10%, si el parámetro estimado resulta del 80%, tendríamos una seguridad del 95% (para $\alpha = 0,05$) de que el parámetro real se sitúa entre el 70% y el 90%. Vemos, por tanto, que la amplitud total del intervalo es el doble del error que introducimos en la fórmula.

Figura 8.2 Tamaño de la muestra Ciudad de La Paz BCS, México.

Muestreo: Ciudad de La Paz, de 18 a 65 años.

Fecha de recogida: 01 de Octubre al 30 de diciembre del 2005.

Universo: 103.811 (INEGI Baja California Sur, 2000).

Nivel de confianza 95 x 100

Margen de posible de error de ± 4 por 100.

Muestra: 524, se desecharon 6 encuestas por errores no muestrales de llenado, quedando para el análisis de datos 518.

$$n = \frac{1,96^2 \times 69 \times 31}{4^2} = \frac{3,92 \times 69 \times 31}{16} = 8384.88 = \underline{524}$$

Figura 8.3 Tamaño de la muestra Ciudad de Barcelona, España.

Muestreo: Ciudad de de Barcelona, España, de 18 a 65 años.

Fecha de recogida: 06 de Febrero al 01 de mayo del 2007

Universo: 1.364.761(INE 2006)

Nivel de confianza 95 x 100

Margen de posible de error de ± 4 por 100.

Muestra: 524.

$$n = \frac{1,96^2 \times 69 \times 31}{4^2} = \frac{3,92 \times 69 \times 31}{16} = 8384.88 = \underline{524}$$

8.5.1 Descripción del área de estudio

Como hemos mencionado anteriormente, este estudio se llevará a cabo en dos ciudades, una de México (La Paz) y otra de España (Barcelona) para hacer la comparación. A continuación se presenta una breve descripción de cada una de ellas, así como las características por las que fueron elegibles.

° La Paz Baja California Sur, México

La ciudad de La Paz, es la capital del estado de Baja California Sur, que se encuentra localizada en la zona noroeste de México. Baja California Sur ocupa la mitad meridional de una larga y esbelta península. Colindando al norte con Baja California Norte, al Este con el Golfo de California y al Oeste con el Océano Pacífico. La Paz esta ubicada en una gran bahía donde el sector de oportunidad son pesca, acuicultura y turismo (Ivanova *et al*, 2002: 391). Como un destino recreativo de renombre, cuenta con vastos recursos naturales y mantiene una atmósfera de ciudad pequeña que la separa de las más desarrolladas zonas de "Mega-Proyectos" que encontramos a lo largo del país.

Tiene una población de 162.954 habitantes y de 103.811 la población de 18 a 65 años (INEGI Baja California Sur, 2000). En el año 2007 tuvo una producción ecológica de 9.541,51 toneladas y su superficie de 890,51 (Oficina Estatal de Información Para el Desarrollo Rural Sustentable (OEIDRUS), Secretaría de Promoción y Desarrollo Económico, Gobierno de Baja California Sur, 2008)

° Barcelona, España

Es la capital de la Comunidad Autónoma de Cataluña y de la provincia homónima, en el noreste de España. Está situada a orillas del mar Mediterráneo, unos 120 Km. al sur de la cadena montañosa de los Pirineos y de la frontera con Francia, en una llanura limitada por el mar, al este, la cordillera litoral (Sierra de Collserola) al oeste, el río Llobregat al sur, y el río Besós al norte. Con una población de 1.605.602 habitantes (INE 2006), Barcelona es la segunda ciudad española más

Capítulo 8 Metodología

poblada y la décima de la Unión Europea. El Área Metropolitana de Barcelona, integrada por 36 municipios, tiene una población de 3.161.081 habitantes y una superficie de 633 km².

Barcelona ha sido escenario de diversos eventos mundiales, que han contribuido a configurar la ciudad y darle proyección internacional. Ha sido durante mucho tiempo una ciudad industrial por excelencia. En la actualidad, las fábricas se han ido desplazando hacia la periferia, lo cual ha permitido reconvertir de forma paulatina los antiguos barrios industriales en nuevas zonas residenciales y de servicio. El puerto de Barcelona se ha convertido en los últimos años en el primer puerto del Mediterráneo en tonelaje de mercancías y en número de contenedores. También es el primer puerto mediterráneo en barcos de cruceros que recalán en la ciudad.

Cuenta con 1.139 operadores de agricultura ecológica y su superficie es de 55.355,00 (Ministerio de Agricultura y Pesca, Secretaría General de Agricultura y Alimentación, Dirección General de Industria Agroalimentaria y Alimentación, Subdirección General de Calidad Agroalimentaria y Agricultura Ecológica, 2006).