



# Capítulo 4

*Escalas Multidimensionales (EMD)*

---



# Capítulo 4

## Escalas Multidimensionales (EMD)

---

### 1. Introducción

---

Las **escalas multidimensionales** (EMD o MDS si se sigue las siglas de su denominación en inglés) engloban a una gran variedad de técnicas multivariantes cuya finalidad es obtener la **estructura** (factores o dimensiones) de los individuos (o variables) subyacente a una matriz de datos empíricos, lo que se consigue al **representar** dicha estructura en una forma geométrica bi o tridimensional. La representación espacial, o “escala” o “espacio métrico”, resultante refleja las propiedades fundamentales de los objetos estudiados al poder establecer una correspondencia entre sus propiedades y sus posiciones en el *continuum* espacial bidimensional o tridimensional de modo que las distancias entre los puntos en el espacio concuerden al máximo con las *disimilitudes* dadas. Este espacio métrico nos permite, pues, identificar proximidades o preferencias entre los objetos con un elevado grado de fidelidad.

Atendiendo a esta doble finalidad, estas técnicas guardan una estrecha relación con las *técnicas factoriales* (a ambas les interesa identificar la estructura interna de los datos, dimensiones o factores, en torno a los cuáles poder resumir y simplificar la matriz de datos originales); y con las *técnicas de agrupamiento o cluster* (ambas comparten el hecho de querer agrupar, en función de su proximidad, a los objetos, personas, estímulos, productos comerciales,... estudiados). Sin embargo, y respecto a su relación con estas técnicas, las EMD tratan de obtener un número muy reducido de dimensiones, idealmente dos o tres (con las técnicas factoriales las dimensiones obtenidas son normalmente más); las EMD trabajan con datos continuos (y no discretos como en el caso de los análisis de cluster) así como con distancias realmente espaciales.

El uso casi generalizado de esta técnica en sociología reside en su capacidad de extraer información métrica a partir de datos ordinales o no métricos. Con demasiada frecuencia los datos con los que se inicia una investigación proceden de las valoraciones

subjetivas, preferencias y juicios de disimilaridad que los entrevistados ofrecen respecto a objetos o conceptos. Otras de las ventajas de las EMD reside en que éstas pueden tratar datos de disimilaridad procedentes de múltiples fuentes, como es el caso de múltiples evaluadores o múltiples sujetos evaluados por un cuestionario.

Un ejemplo puede simplificar lo expuesto. La denominación de “Coche del año” parte, precisamente, de la aplicación de las EMD. ¿Cómo percibe el público las diferencias entre distintos coches? Si disponemos de datos de valoraciones (de similaridad) emitidas por los sujetos sobre las diferentes marcas y modelos de coches, se puede utilizar el escalamiento multidimensional para identificar las dimensiones que describan las preferencias de los consumidores. El precio y el tamaño de los vehículos suelen ser las dos dimensiones que explican las similitudes de las que informan los encuestados y, en consecuencia, en torno a estos dos conceptos establecen sus jerarquía de preferencias.

Las EMD se apoyan en lo que se conoce como **criterio de orden débil**. Este criterio establece que si la semejanza entre un par de “objetos” ( $i$  y  $j$ ) de un conjunto de  $n$ , es mayor que la semejanza entre otro par de “objetos” ( $k$  y  $l$ ) a nivel gráfico la distancia entre  $i$  y  $j$  será menor que la distancia entre  $k$  y  $l$ . Lo que se busca es encontrar un conjunto de puntos en un espacio de dimensión mínima tales que las distancias entre los pares de puntos,  $d_{ij}$ , (datos de disimilaridad) estén monotónicamente relacionados con los datos que miden la proximidad de los datos en el sentido de que:

$$d_{ij} < d_{kl} \text{ para todo } s_{ij} > s_{kl}$$

En este capítulo vamos a ver:

1. Estadísticos para valorar la bondad de ajuste de los datos.
2. El proceso para llegar a la obtención de la dimensionalidad subyacente.
3. La representación gráfica o métrica de los datos sobre el espacio conceptual (normalmente de dos o tres dimensiones).
4. Interpretación del espacio métrico, o escala, obtenido.

## 2. Estimación de la bondad de ajuste de los datos al modelo

Antes de determinar las dimensiones o estructura subyacente de los datos debemos valorar si los datos o configuración de puntos se ajustan a un modelo de EMD y, en consecuencia, puede ser valorado a partir de sus consideraciones. Al respecto los indicadores con los que contamos y que recoge la salida de resultados del del SPSS son:

### a.- *S-stress de Young.*

La s-stress se define en términos de distancias al cuadrado. El programa trabaja generando sucesivas iteraciones hasta conseguir un s-stress bajo pues podemos afirmar que se tiene un buen ajuste cuando estos valores son bajos. Como tendremos ocasión de apreciar, a medida que el número de iteraciones es mayor, el s-stress es cada vez más pequeño. Las iteraciones se paralizan cuando la disminución del s-stress es inferior a 0.001.

### b.- *Stress de Kruskal.*

Por su parte, el stress de Kruskal se define en términos de distancias. Este indicador se expresa con la letra S y adquiere un valor que oscila entre 0 y 1. Cuanto menor es el stress, mejor es el ajuste. El inconveniente de este indicador reside en el hecho de que no existen unos valores límites que permitan decidir si el ajuste es bueno o no, por lo que dependerá del número de estímulos. Para facilitar su interpretación y valoración contamos con Tablas que establecen el valor mínimo en función de los estímulos (propuestas por el propio Kruskal y Spence y Ogilve). A partir de éstas, cuanto más estímulos tengamos, más bajo deberá ser este valor y, viceversa.

### c.- *Correlación múltiple al cuadrado (RSQ).*

El valor que resulta de la correlación múltiple al cuadrado entre los puntos escalares estimados y los actuales, puede interpretarse como la proporción de varianza explicada por las distancias respectivas. Se tiene un buen ajuste cuando la correlación múltiple al cuadrado es alta.

### d.- *Diagrama de Shepard o gráfico de ajuste lineal.*

El diagrama de Shepard nos ofrece la posibilidad de detectar cualquier anomalía. La configuración estará bien ajustada cuan-

do el diagrama refleje una función creciente. Un buen ajuste se representa con una diagonal en el citado gráfico.

### 3. Obtención de la dimensionalidad

En EMD la decisión acerca del número de factores a conservar depende del marco teórico con el que se haya iniciado la aplicación de esta técnica y, en consecuencia, del conocimiento que el propio investigador tenga acerca del propio tema objeto de análisis.

Como norma general, y en lo que respecta a las EMD (este no es el caso en el análisis factorial), nos ceñiremos al menor número de dimensiones, lo que implica que en número no excederán de tres: a menor número de dimensiones más fácil la interpretación; aunque también es cierto que a mayor número de dimensiones mejor es el ajuste.

Pese a que la pericia y conocimiento del investigador sea en este caso determinante, podremos utilizar los estadísticos s-stress y stress como criterios a la hora de delimitar el número de dimensiones. Por su parte, la representación gráfica de la nube de puntos también nos puede servir para este propósito.

### 4. Representación gráfica de los datos e interpretación

Una de las ventajas que introducen las EMD es la posibilidad de representar gráficamente los resultados. La comprensión e interpretación de las dimensiones obtenidas se simplifica al ubicar los puntos en un espacio geométrico. La lectura del gráfico solo es posible cuando en éste aparecen no más de tres dimensiones. Ya hemos comentado como se dificulta su interpretación a medida que vamos incrementando el número de dimensiones.

Si la gráfica bidimensional o tridimensional es la mejor salida de los análisis efectuados aplicando EMD, nuestro interés debe centrarse en la interpretación de los ejes o direcciones que enmarca la representación espacial. Nuevamente, ante esta circunstancia, el conocimiento que el investigador tenga sobre el tema es de suma importancia. Sin embargo, podemos tener en cuenta que en la medida que nos alejamos en el espacio siguiendo una dirección particular, los puntos que se van encontrando

corresponden a objetos que poseen más de una particular e identificable propiedad, pudiendo concluir en que la citada propiedad en importante es los procesos que dan lugar a los datos.

No debemos olvidar que el objetivo de esta técnica de investigación es el de identificar las dimensiones subyacentes para, posteriormente, ordenar de mayor a menor los elementos asociados a cada dimensión según la escala y la puntuación factorial obtenida.

### 5. Cuadro de Diálogo de las Escalas Multidimensionales

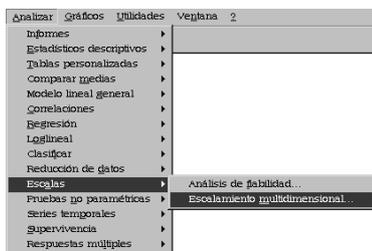


Figura 1

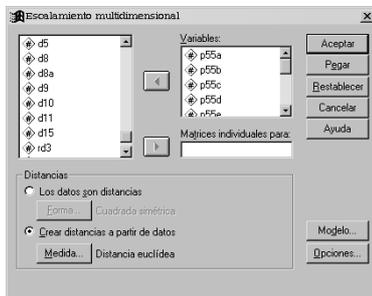


Figura 2

Una vez ya comentados los aspectos básicos y fundamentales que tenemos que tener en cuenta a la hora de abordar un análisis con EMD podemos profundizar más en las especificaciones a partir de su correspondiente cuadro de diálogo.

El análisis, en sentido estricto, empieza una vez que hemos accedido al Cuadro de Diálogo de escalamiento Multidimensional

**1º paso:** Para acceder a él deberemos seguir la secuencia **Analyze: Scale: Escalamiento multidimensional** (figura 1).

**2º paso:** Una vez en el cuadro de diálogo principal de EMD seleccionaremos del listado de variables que forman la matriz de datos originales, aquellas variables con las que vamos a realizar el análisis. Seleccionadas las variables, las trasladaremos a la ventana de **Variables** (figura 2).

En nuestro caso particular, las variables seleccionadas, corresponden a un conjunto de variables procedentes de una misma pregunta del cuestionario del CIRES. Nuestro interés reside en agrupar las variables de la pregunta P55 en la que se pedía a los encuestados su posicionamiento, dentro de una escala, sobre cuestiones relacionadas con aptitudes discriminatorias (los NS/NC han sido suprimidos del análisis). Vamos a tratar de agruparlas en una serie de dimensiones que luego nos permitirán determinar su jerarquía (de mayor a menor) dentro de la dimensión a la que pertenezcan.

**3º paso:** Una vez seleccionadas las variables, debemos especificar si los datos de las variables seleccionadas son datos de disi-

milaridades (contamos con una matriz de datos de distancias); o si bien, de lo contrario partimos de una matriz de datos multidimensionales. Si nuestro caso fuera el primero seleccionaríamos de la opción de **Distancias** que figura en la parte inferior de cuadro de diálogo principal: **Los datos son distancias**; y si es el segundo, la opción a señalar sería la que se nombra como **Crear distancias a partir de datos**.

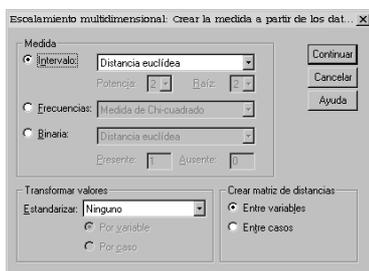


Figura 3

Debemos recordar que las EMD generan sus análisis a partir de una matriz de distancias y nunca a partir de datos reales. Sin embargo, lo más frecuente es disponer de los datos reales y no de las distancias. Por ello, antes de proceder a los distintos análisis, transformemos los datos originales en distancias. La opción que nos lo permite es la segunda de las arriba indicadas y al clicar sobre ella accedemos al subcuadro de diálogo **Crear la medida a partir de los datos**. En él nos encontramos con distintos niveles de medida en función del tipo de datos que dispongamos. En concreto, éstas son: de **Intervalo**, **Frecuencias** y **Binaria**. En cada una de estas tres opciones parece una lista desplegable de medidas de la que deberemos seleccionar una de ellas. Al seleccionar una de ellas, y si regresamos al cuadro principal de las EMD, veremos como la opción escogida aparece junto al botón de comando de **Medida**. En nuestro caso particular debemos asegurarnos de que la opción marcada es **Cálculo de las distancias a partir de la medida de distancia euclídea** (figura 3). Esta es la opción que viene por defecto cuando el tipo de medida es la de intervalo.

Cabe recordar que es importante elegir el nivel de medida más indicado pues nos va asegurar que el escalamiento que se realice sobre las variables sea el óptimo. En última instancia, el buen ajuste de un modelo de los datos puede depender de la medida de proximidad elegida.

En determinados casos, y en concreto en aquellos en los que las variables seleccionadas están medidas en escalas muy diferentes (en pesetas, kilómetros,...), resulta más indicado **Transformar los valores** (tipificarlos) antes de calcular las distancias (esta opción no está disponible para datos binarios). Por defecto, y en el subcuadro de diálogo Crear la medida a partir de los datos, aparece la opción **Ninguna**. En el caso de que sí se quieran transformar los datos habrá que seleccionar un método de tipificación de la lista desplegable **Estandarizar**.

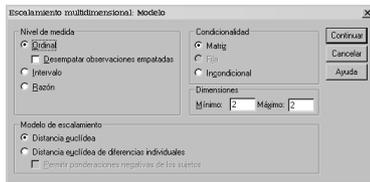


Figura 4

**4º paso:** Una vez que ya contamos con una matriz de distancias y no de datos reales, debemos solicitar al programa que genere a partir de ésta el modelo. Para ello accedemos al subcuadro de diálogo de **Modelo** cliqueando uno de los dos botones de comando que aparecen en la derecha del cuadro de diálogo principal de EMD. Allí seleccionaremos: el **Nivel de medida** (que en nuestro caso concreto debe ser el ordinal); el **Modelo de escalamiento** (normalmente distancia euclídea); y, el número máximo y mínimo de **Dimensiones** (que como ya sabemos no debe exceder de 3) (figura 4).

En función de las especificaciones introducidas en este cuadro de diálogo, el programa calculará los distintos estadísticos vinculados a la valoración de bondad de ajuste de las variables, a saber: **s-stress** (en el ejemplo que recogemos en la sección de Resultados y tras las especificaciones mencionadas, el programa ha utilizado 9 iteraciones hasta alcanzar un s-stress de 0.00100); **stress** (en nuestro ejemplo con 9 estímulos, dos dimensiones, un stress de 0.05181 y teniendo en cuenta las tablas que antes mencionamos, podemos considerar que el ajuste es bueno); y, **RQS** (que al arrojar un valor alto, 0.99006, nos permite concluir que el ajuste alcanzado es bueno).

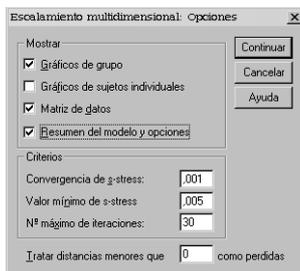


Figura 5

Determinados **Criterios** como: el **Número máximo de iteraciones** (30); el valor de **Convergencia de s-stress** (0.001); y el **Valor mínimo de s-stress** (0.005); así como el tipo de salida con las que queremos que nos **Muestre** el programa los resultados arrojados por el análisis (Gráficos y/o Matrices), se debe concretar en el subcuadro de diálogo de **Opciones** que aparece una vez que hemos cliqueado en el citado botón de comando en el cuadro de diálogo principal de las EMD (figura 5).

**5º paso:** Ratificado el ajuste del modelo, pasamos a interpretar gráficamente los datos obtenidos en la dimensionalidad de las variables. En función de las dimensiones obtenidas y las puntuaciones obtenidas por cada variable en cada dimensión, volvemos a calcular la matriz de distancias y establecemos gráficamente en una escala, en nuestro ejemplo de dos dimensiones, la posición de cada una de las variables.

## 6. Bibliografía Comentada

---

La aplicación de las EMD en las Ciencias Sociales se encuentra muy generalizada. Buena muestra de ello es el importante volumen de investigaciones, que desde distintas áreas de conocimiento, han aplicado esta técnica en la resolución de sus objetivos. Nosotros recogemos tres de ellas.

- Bigne, J. Enrique y Vila, Natalia (1998): “Eficacia de la técnica de EMD en la medición del posicionamiento: una aplicación en los servicios”, *ESIC Market*, nº 100, pp. 71-90.

*En este artículo se ratifica la idoneidad de las EMD como técnica que permite visualizar la posición relativa de una oferta de servicios frente a los competidores una vez adoptado el posicionamiento elegido. Se hace hincapié en que las EMD representan en un mapa las posiciones de distintas ofertas competidoras cuya proximidad estará en función del grado en que compitan.*

- Colodrón, María y Montorio, Ignacio (1998): “La imagen de las personas mayores mediante una tarea de asociación de atributos en familiares cuidadores”, *Revista de psicología social*, nº 13, pp. 389-398.

*Este trabajo estudia la relación entre las imágenes que los cuidadores informales mantienen sobre el grupo de personas mayores en general y sobre el familiar mayor al quien cuidan en particular. Una muestra de 95 sujetos agrupó 30 adjetivos según un criterio espontáneo de similitud (estos adjetivos habían sido obtenidos a partir de los utilizados por otra muestra de 100 sujetos que describieron a personas con distintas características). Mediante el escalamiento multidimensional de la matriz de proximidades se obtuvo una solución tridimensional que diferenció los adjetivos en cuatro grupos. Posteriormente, una muestra de 62 cuidadores seleccionaron de entre los 30 adjetivos aquellos con los que identificaban a una persona mayor, en general, y con los que identificaban a su familiar mayor. Los resultados muestran que mientras los cuidadores muestran un mismo número de características positivas para describir a las personas mayores y a su familiar, eligen un*

*mayor número de características negativas para describir al grupo de personas mayores.*

- García Ricardo, Arce Constantino y Sabucedo Jose Manuel (1997): “Evaluación de la preferencia ambiental mediante escalamiento multidimensional”, *Revista de psicología social aplicada*, n° 7, pp. 53-75.

*Esta investigación realiza una evaluación sobre las preferencias ambientales que motivan a los ciudadanos a elegir barrio en el momento de fijar su residencia. Los resultados indican que las dimensiones más importantes son: las relacionadas con la posición social (ligado a la centralidad del barrio), el mantenimiento y conservación del barrio (ligado a las características de seguridad y actividad) y la espaciosidad del barrio (ligado al volumen construido).*

## 7. Resultados

---

Los resultados que aparecen a continuación hacen referencia a la aplicación de las escalas multidimensionales en la pregunta que recoge la posición de los encuestados, dentro de una escala, respecto a nueve cuestiones relacionadas con la discriminación. El interés reside, pues, en identificar las dimensiones (o grupos de variables) implícitas en la citada pregunta para, y a partir de sus respectivas puntuaciones factoriales y según la escala de cada dimensión, ordenar de mayor a menor cada uno de los ítems asociados a cada una de las dimensiones identificadas.

- Una vez anotadas todas las especificaciones señaladas el programa presenta dos tablas con carácter exclusivamente informativo. La primera de ellas hace un resumen global de los datos implicados en el análisis; mientras que la segunda hace lo propio respecto a cada una de las variables seleccionadas.
- El análisis, propiamente, empieza con la matriz de distancias. Al respecto cabe recordar que el cálculo de esta matriz es fundamental pues: las EMD solo trabajan a partir de éstas y las EMD definen las dimensiones a partir de la proximidad que manifiesten los datos.

- Es necesario someter los datos del modelo, en este caso la matriz de distancias, a pruebas de bondad de ajuste de ahí que las siguientes tablas que se presentan son las de las pruebas de S-stress, stress y RSQ.
- Una vez ratificado la bondad de los datos el programa nos ofrece las dimensiones (no más de tres) con las puntuaciones factoriales asociadas a cada una de las variables en cada una de las dimensiones. A partir de estas puntuaciones podemos saber: qué variables pertenecen a cada una de las dimensiones (una variable se adscribe a la dimensión en la que presente mayor puntuación); y qué orden mantienen las variables adscritas a una dimensión (una vez determinado las variables que forman cada una de las dimensiones, en cada dimensión debemos ordenarlas de mayor (valores positivos) a menor (valores negativos)).
- A partir de las dimensiones se obtiene una nueva matriz de distancias. Nótese como sus valores aparecen más ajustados que en la matriz de distancias originales.
- La presentación de resultados concluye con la exposición de una serie de gráficos.

### 7.1. Resumen del procedimiento

Casos					
Valid		Perdidos		Total	
N	Porcentaje	N	Porcentaje	N	Porcentaje
891	74,3%	309	25,8%	1200	100,0%

<sup>a</sup>. Distancia euclídea usada

```
Data Options
Number of Rows (Observations/Matrix).....9
Number of Columns (Variables).....9
Number of Matrices .....1
Measurement Level .....Ordinal
Data Matrix Shape.....Symmetric
Type .....Dissimilarity
Approach to Ties .....Leave Tied
Conditionality .....Matrix
Data Cutoff at .....0,000000
```

```

Model Options

Model ..... Euclid
Maximum Dimensionality ..... 2
Minimum Dimensionality ..... 2
Negative Weights ..... Not Permitted

Output Options

Job Option Header ..... Printed
Data Matrices ..... Printed
Configurations and Transformations ..... Plotted
Output Dataset ..... Not Created
Initial Stimulus Coordinates ..... Computed

Algorithmic Options

Maximum Iterations ..... 30
Convergence Criterion ..... ,00100
Minimum S-stress ..... ,00500
Missing Data Estimated by ..... Ulbounds
Tiestore..... 36
    
```

## 7.2. Matriz de Distancias

Raw (unscaled) Data for Subject 1

	1	2	3	4	5
1	,000				
2	45,837	,000			
3	49,092	53,879	,000		
4	49,840	55,127	34,699	,000	
5	45,067	36,579	51,971	52,564	,000
6	61,498	63,253	63,151	59,850	63,427
7	46,648	42,485	49,417	49,598	44,193
8	62,618	62,498	57,088	54,397	63,214
9	47,434	44,710	58,992	58,292	47,064

	6	7	8	9
6	,000			
7	63,953	,000		
8	46,551	64,008	,000	
9	63,624	44,068	65,276	,000

### 7.3. Bondad de Ajuste de las Variables (1ª prueba)

Iteration history for the 2 dimensional solution (in squared distances)

Young's S-stress formula 1 is used.

Iteration	S-stress	Improvement
1	,05047	
2	,03673	,01374
3	,03041	,00633
4	,02644	,00396
5	,02387	,00258
6	,02221	,00166
7	,02088	,00133
8	,01982	,00106
9	,01905	,00076

Iterations stopped because

S-stress improvement is less than ,001000

### • Bondad de Ajustes de las Variables (2ª y 3ª prueba)

Stress and squared correlation (RSQ) in distances

RSQ values are the proportion of variance of the scaled data (disparities) in the partition (row, matrix, or entire data) which is accounted for by their corresponding distances. Stress values are Kruskal's stress formula 1.

For matrix

Stress = ,05181      RSQ = ,99006

### 7.4. Identificación de Dimensiones y Puntaciones Factoriales

Configuration derived in 2 dimensions

Stimulus Number	Stimulus Name	Dimension	
		1	2
1	P55A	,8549	,0761
2	P55B	,8840	,4739
3	P55C	-,0846	-1,4147
4	P55D	-,4075	-1,2406
5	P55E	,8711	,2357
6	P55F	-2,0213	,7994
7	P55G	,9463	,0795
8	P55H	-2,0771	,1579
9	P55I	1,0342	,8328

### 7.5. Reajuste de la Matriz de Distancias en función de las dimensiones

Optimally scaled data (disparities) for subject 1

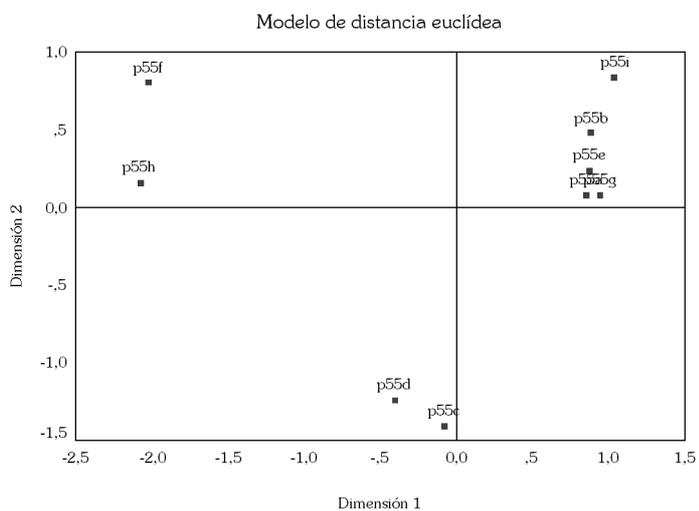
	1	2	3	4	5
1	,000				
2	,434	,000			
3	1,762	2,122	,000		
4	1,858	2,162	,309	,000	
5	,434	,309	1,907	1,953	,000
6	2,948	2,948	2,948	2,601	2,948
7	,460	,399	1,815	1,858	,434
8	2,948	2,948	2,525	2,162	2,948
9	,778	,434	2,525	2,525	,619

	6	7	8	9
6	,000			
7	3,045	,000		
8	,460	3,045	,000	
9	3,045	,434	3,184	,000

### 7.6. Representación gráfica de las dimensiones

Configuración de estímulos derivada



### 7.7. Representación gráfica de Ajuste Lineal

