

Modelos Loglineales

- Trata de interpretar los efectos que sobre las frecuencias observadas entre las categorías de varias variables categóricas tienen cada una de ellas así como sus interrelaciones.

➡ Para ello trabaja con el logaritmo natural (neperiano) de las frecuencias obtenidas en el cruce de varias variables

➡ Imaginemos una tabla sencilla de sencilla de dos variables sexo y grado de felicidad percibido cuyas frecuencias son:

	Nada Feliz	Poco Feliz	Bastante F.	Muy Feliz	Total
Hombre	3	75	449	50	577
Mujer	13	90	461	52	616
Total	16	165	910	102	1193

➡ La tabla convertida en sus logaritmos neperianos queda como sigue:

	Nada Feliz	Poco Feliz	Bastante F.	Muy Feliz	Media
Hombre	1,099	4,317	6,107	3,912	3,859
Mujer	2,565	4,500	6,133	3,951	4,287
Media	1,832	4,409	6,120	3,932	4,073

- En el modelo loglineal se cumple que para cualquier casilla ij el logaritmo natural de su frecuencia es:

$$\ln(n_{ij}) = \mu + \lambda_i^f + \lambda_j^c + \lambda_{ij}^{fc}$$

➡ Donde :

μ = Es la media de los logaritmos de todas las celdas de la tabla

λ_i^f = El efecto de la categoría i de la variable F

λ_j^c = El efecto de la categoría j de la variable C

λ_{ij}^{fc} = El efecto de la interacción «i» y «j» de las dos de la variables

➡ De esta forma, el logaritmo natural de la primera casilla es $\ln(n_{11}) = 1,099$

$$\mu = 4,0731$$

$$\lambda_1^f = 3,8588 - 4,0731 = -0,2143$$

$$\lambda_1^c = 1,83318 - 4,0731 = -2,2413$$

$$\lambda_{11}^{fc} = 1,0986 - (4,0731 - (0,2143 + 2,2413)) = -0,51189$$

➡ Estimados así todos los parámetros λ_{ij} de nuestra tabla:

	Nada Feliz	Poco Feliz	Bastante F.	Muy Feliz	Total
Hombre	-0,5189	0,1231	0,2011	0,1947	-0,2143
Mujer	0,5188	-0,1232	-0,2011	-0,1947	0,2143
Total	-2,2413	0,3356	2,0471	-0,1415	4,0731

- La suma de los parámetros λ en cada factor (marginales) es igual a cero. Los valores λ positivos o negativos corresponden a frecuencias por encima o debajo de las medias. Así un λ negativo para el factor hombre de -0,2143 significa que hay más hombres que mujeres.
 - ➔ En cuanto al λ_{i11}^{cf} correspondiente a la categoría hombre y sentirse «Nada feliz» -0,5189 indica que tienen menor tendencia que las mujeres a sentirse así.
- Este modelo donde se contemplan todos los efectos se denomina «**Modelo saturado general**» y nos permite explicar la frecuencia de cualquier casilla siguiendo la fórmula señalada

$$\ln(n_{ij}) = \mu + \lambda_i^f + \lambda_j^c + \lambda_{ij}^{cf} \quad \text{Caso de dos variables } f \text{ y } c$$

Con tres variables f , c y d sería:

$$\ln(n_{ijk}) = \mu + \lambda_i^f + \lambda_j^c + \lambda_k^d + \lambda_{ij}^{cf} + \lambda_{ik}^{cd} + \lambda_{jk}^{fd} + \lambda_{ijk}^{cfd}$$

En general, con t variables sería:

$$\ln(n_{i_1 i_2 i_3 \dots i_t}) = \mu + \lambda_{i_1}^{x_1} + \lambda_{i_2}^{x_2} + \dots + \lambda_{i_t}^{x_t} + \lambda_{i_1 i_2}^{x_1 x_2} + \dots + \lambda_{i_{t-1} i_t}^{x_{t-1} x_t} + \dots + \lambda_{i_1 i_2 i_3}^{x_1 x_2 x_3} + \dots + \lambda_{i_1 i_2 \dots i_t}^{x_1 x_2 \dots x_t}$$

Modelos Loglineales. El modelo saturado

- Por tanto, se denomina «**Modelo saturado**» aquel que para su cálculo contempla todos los posibles efectos, es decir si trabajamos con las tres variables A, B y C tendrá en cuenta:

$$M = \mu + A + B + C + (A * B) + (A * C) + (B * C) + (A * B * C)$$

➔ μ = El efecto global de la distribución en la tabla de frecuencias

➔ $\left. \begin{matrix} A \\ B \\ C \end{matrix} \right\}$ Los efectos individuales de cada variable

➔ $\left. \begin{matrix} A * B \\ A * C \\ B * C \end{matrix} \right\}$ Los efectos de interacción de segundo orden entre las variables

➔ $A * B * C$ Los efectos de interacción de tercer orden entre las variables

- El modelo saturado permite reproducir con exactitud la frecuencia de cualquier casilla

Modelos Loglineales. Modelización

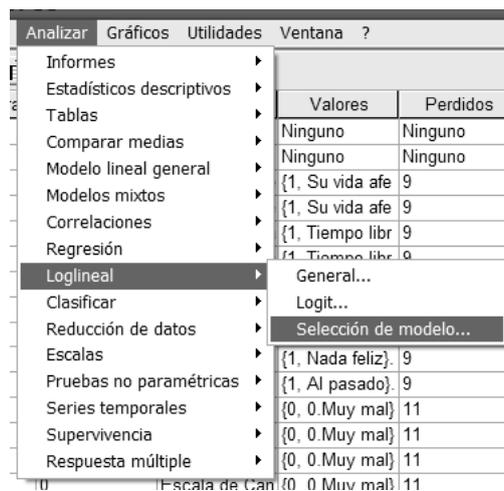
- Un caso especial es el modelo donde no se contemplan los efectos debidos a la interacción de las distintas variables, es decir solo se contemplan los efectos de las variables. Se le denomina «**Modelo de interdependencia**»

$$M = \mu + A + B + C \quad \ln(n_{i_1 i_2 i_3 \dots i_r}) = \mu + \lambda_{i_1}^{x_1} + \lambda_{i_2}^{x_2} + \dots + \lambda_{i_r}^{x_r}$$

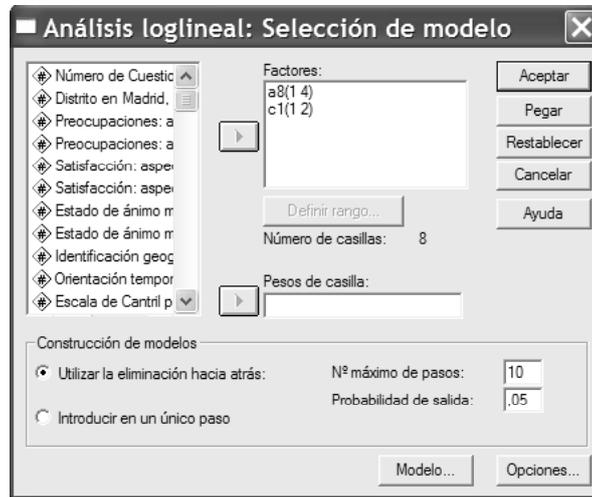
- La modelización loglineal calcula el modelo saturado teniendo en cuenta todos los efectos. A partir del cual va eliminando efectos y contrastando si las frecuencias obtenidas al suprimir este varían o no significativamente de las observadas (que coinciden con el modelo saturado)
 - ➡ Si las diferencias encontradas al eliminar un efecto no difieren significativamente de las observadas, quiere decir que dicho efecto puede suprimirse del modelo ya que no contribuye significativamente a explicar las frecuencias encontradas en el cruce de estas variables
 - ➡ El programa continuará en sucesivas interacciones eliminando efectos hasta dar con el modelo más simple que explique las frecuencias observadas

Modelización, con SPSS

- En el SPSS vamos a «**Analizar → loglineal → Selección del modelo**»

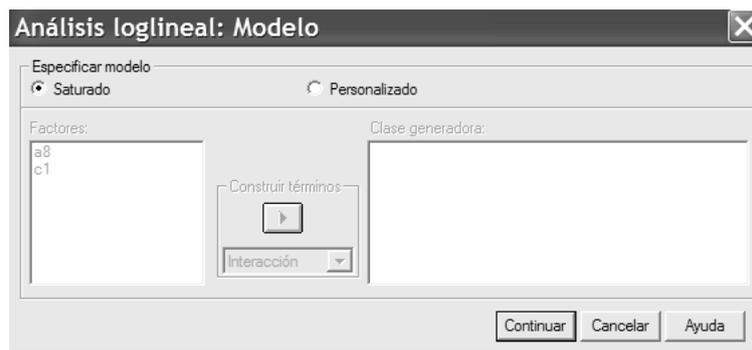


- En el menú emergente introducimos las variables a analizar seso [C1] y grado de felicidad [a8] como factores



- ➔ Se definen los rangos de las variables a utilizar seleccionando como criterio de construcción «*Eliminación hacia atrás*» de forma que cada factor que entra en el análisis es testado saliendo del modelo si se cumple con la probabilidad superior al 0,05 en el ratio de máxima verosimilitud

- En el botón «*Modelo*» solicitamos saturado que viene por defecto

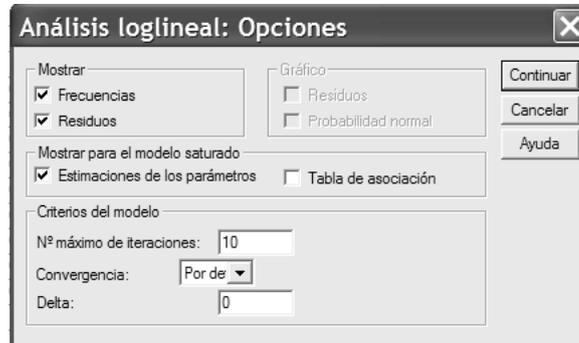


- ➔ El modelo saturado es aquel donde intervienen todos los efectos posibles. En nuestro caso con dos variables «*c*» y «*f*» los efectos serían el global de la tabla, los individuales de las variables, mas el de interacción entre ellos es decir: $\mu + \lambda^c_i + \lambda^f_j + \lambda^{cf}_{ij}$ simplificando la anotación $\mu + c + f + (c*f)$

- ➔ Con tres variables «*c*» «*x*» y «*f*» sería :

$$\mu + \lambda^c_i + \lambda^x_k + \lambda^f_j + \lambda^{cf}_{ij} + \lambda^{cx}_{ik} + \lambda^{xf}_{kj} + \lambda^{cxf}_{ijk}$$
simplificando la anotación $\mu + c + x + f + (c*x) + (c*f) + (x*f) + (c*x*f)$

- En opciones solicitamos «*Frecuencias*», «*Residuos*» y «*Estimaciones de los parámetros*»



- ➔ Como criterios del modelo dejamos los valores por defecto: «*Nº de interacciones*» [10], «*Convergencia*» [Por defecto] y «*Delta*»[0]

- La salida del SPSS nos muestra en primer lugar los casos válidos y computados (de los 1193 casos hay 7 con valores nulos en alguna de las variables). Así como las categorías que contiene cada variable del análisis

		N
Casos	Válido	1193
	Fuera del rango ^a	0
	Perdido	7
	Válido ponderado	1193
Categorías	a8	4
	c1	2

a. Casos rechazados debido a valores de los factores fuera del rango.

- A continuación nos informa de los factores que entran en el modelo en nuestro caso la interacción de las variables «a8» y «c1» ya que es un modelo saturado y entran todas (las de nivel 2 a8*c1 y las de nivel uno a8 y C1). Cerrando el modelo en la interacción 1 al cumplirse los criterios de convergencia especificados

Información sobre la convergencia

Clase generadora	a8*c1	
Número de iteraciones		1
Diferencia máxima entre observados y marginales ajustados		,000
Criterio de convergencia		,461

- Seguidamente, nos muestra la tabla de frecuencias bajo el modelo saturado, dicha tabla coincide con los valores observados ya que contempla los efectos de todas las variables y sus posibles interacciones

Frecuencias y residuos de casillas

a8	c1	Observado		Esperado		Residuos	Residuos típicos
		Recuento ^a	%	Recuento	%		
Nada feliz	Hombre	3,500	,3%	3,500	,3%	,000	,000
	Mujer	13,500	1,1%	13,500	1,1%	,000	,000
Poco feliz	Hombre	75,500	6,3%	75,500	6,3%	,000	,000
	Mujer	90,500	7,6%	90,500	7,6%	,000	,000
Bastante feliz	Hombre	449,500	37,7%	449,500	37,7%	,000	,000
	Mujer	461,500	38,7%	461,500	38,7%	,000	,000
Muy feliz	Hombre	50,500	4,2%	50,500	4,2%	,000	,000
	Mujer	52,500	4,4%	52,500	4,4%	,000	,000

a. Para modelos saturados, se ha añadido ,500 a todas las casillas observadas.

- El SPSS nos informa que realiza en interacciones sucesivas la eliminación de efectos comenzando por los de mayor nivel . Si las frecuencias calculadas sin él no difieren significativamente (utiliza el chi-cuadrado) de las observadas lo elimina y pasa a examinar el siguiente efecto

Contrastes de bondad de ajuste

	Chi-cuadrado	gl	Sig.
Razón de verosimilitudes	,000	0	.
Pearson	,000	0	.

Efectos de orden K y superior

	K	gl	Razón de verosimilitudes		Pearson		Número de iteraciones
			Chi-cuadrado	Sig.	Chi-cuadrado	Sig.	
Efectos de orden K superiores ^a	1	7	1530,684	,000	1712,140	,000	0
	2	3	7,026	,071	6,543	,088	2
Efectos de orden K ^b	1	4	1523,658	,000	1705,597	,000	0
	2	3	7,026	,071	6,543	,088	0

a. Contrasta que los efectos de orden k superiores son cero.

b. Contrasta que los efectos de orden k son cero.

- En la siguiente tabla nos muestra las interacciones realizadas y los resultados de las pruebas de significación

Contrastes de bondad de ajuste

	Chi-cuadrado	gl	Sig.
Razón de verosimilitudes	,000	0	.
Pearson	,000	0	.

Efectos de orden K y superior

	K	gl	Razón de verosimilitudes		Pearson		Número de iteraciones
			Chi-cuadrado	Sig.	Chi-cuadrado	Sig.	
Efectos de orden K superiores ^a	1	7	1530,684	,000	1712,140	,000	0
	2	3	7,026	,071	6,543	,088	2
Efectos de orden K ^b	1	4	1523,658	,000	1705,597	,000	0
	2	3	7,026	,071	6,543	,088	0

a. Contrasta que los efectos de orden k superiores son cero.

b. Contrasta que los efectos de orden k son cero.

- La siguiente tabla resume los resultados de las interacciones realizadas. En la interacción 1 se eliminó el efecto combinado de $a\delta*c1$ dado que su eliminación no alteraba significativamente las frecuencias calculadas

Estadísticos de eliminación hacia atrás

Resumen de los pasos

Paso ^b	Efectos	Chi-cuadrado ^a	gl	Sig.	Número de iteraciones
0	Clase generadora ^c	$a\delta*c1$,000		
	Efecto eliminado 1	$a\delta*c1$	7,026	,071	2
1	Clase generadora ^c	$a\delta, c1$	7,026	,071	
	Efecto eliminado 1	$a\delta$	1522,382	,000	2
	2	$c1$	1,275	,259	2
2	Clase generadora ^c	$a\delta$	8,301	,081	
	Efecto eliminado 1	$a\delta$	1522,383	,000	0
3	Clase generadora ^c	$a\delta$	8,301	,081	

a. Para 'efecto eliminado', éste es el cambio en la Chi-cuadrado después de eliminar el efecto del modelo.

b. En cada paso, se elimina el efecto con mayor nivel de significación para el Cambio en la razón de verosimilitudes, siempre que el nivel de significación sea mayor que ,050.

c. Se muestran los estadísticos para el mejor modelo en cada paso desoués del paso 0.



En la interacción 2 se elimina el efecto de la distribución de la variable $c1$ quedando explicadas las frecuencias en base a la distribución de la variable $a\delta$ y terminando en este paso el proceso

- Finalmente nos da la tabla de frecuencias calculada bajo el modelo resultante y la prueba de significación. En este caso, el modelo se puede explicar por la distribución de $a\delta$ sin que las frecuencias resulten significativamente diferentes.

Frecuencias y residuos de casillas

a δ	c1	Observado		Esperado		Residuos	Residuos típicos
		Recuento	%	Recuento	%		
Nada feliz	Hombre	3,000	,3%	8,000	,7%	-5,000	-1,768
	Mujer	13,000	1,1%	8,000	,7%	5,000	1,768
Poco feliz	Hombre	75,000	6,3%	82,500	6,9%	-7,500	-,826
	Mujer	90,000	7,5%	82,500	6,9%	7,500	,826
Bastante feliz	Hombre	449,000	37,6%	455,000	38,1%	-6,000	-,281
	Mujer	461,000	38,6%	455,000	38,1%	6,000	,281
Muy feliz	Hombre	50,000	4,2%	51,000	4,3%	-1,000	-,140
	Mujer	52,000	4,4%	51,000	4,3%	1,000	,140

Contrastes de bondad de ajuste

	Chi-cuadrado	gl	Sig.
Razón de verosimilitudes	8,301	4	,081
Pearson	7,811	4	,099