

Estadística IV

28/08/2024

Segundo Semestre 2024

Profesor: Kevin Carrasco

Daniela Olivares

Ayudante: María Fernanda Núñez

Universidad Alberto Hurtado

Departamento de Sociología

Análisis de variables observadas categóricas

Análisis de Correspondencia simple y múltiple

*****Bibliografía recomendada*****

- Greenacre, Michael. 2008. La práctica del análisis de correspondencias. Madrid: Fundación BBVA. Caps 1 - 10
- Vivanco, M. (1999): Análisis estadístico multivariable. Ed. Universitaria. Caps. 5 y 6.
- Hair J, F. et al (2004): *Análisis multivariante*. Madrid: Pearson. Cap. 10.

Análisis de correspondencias

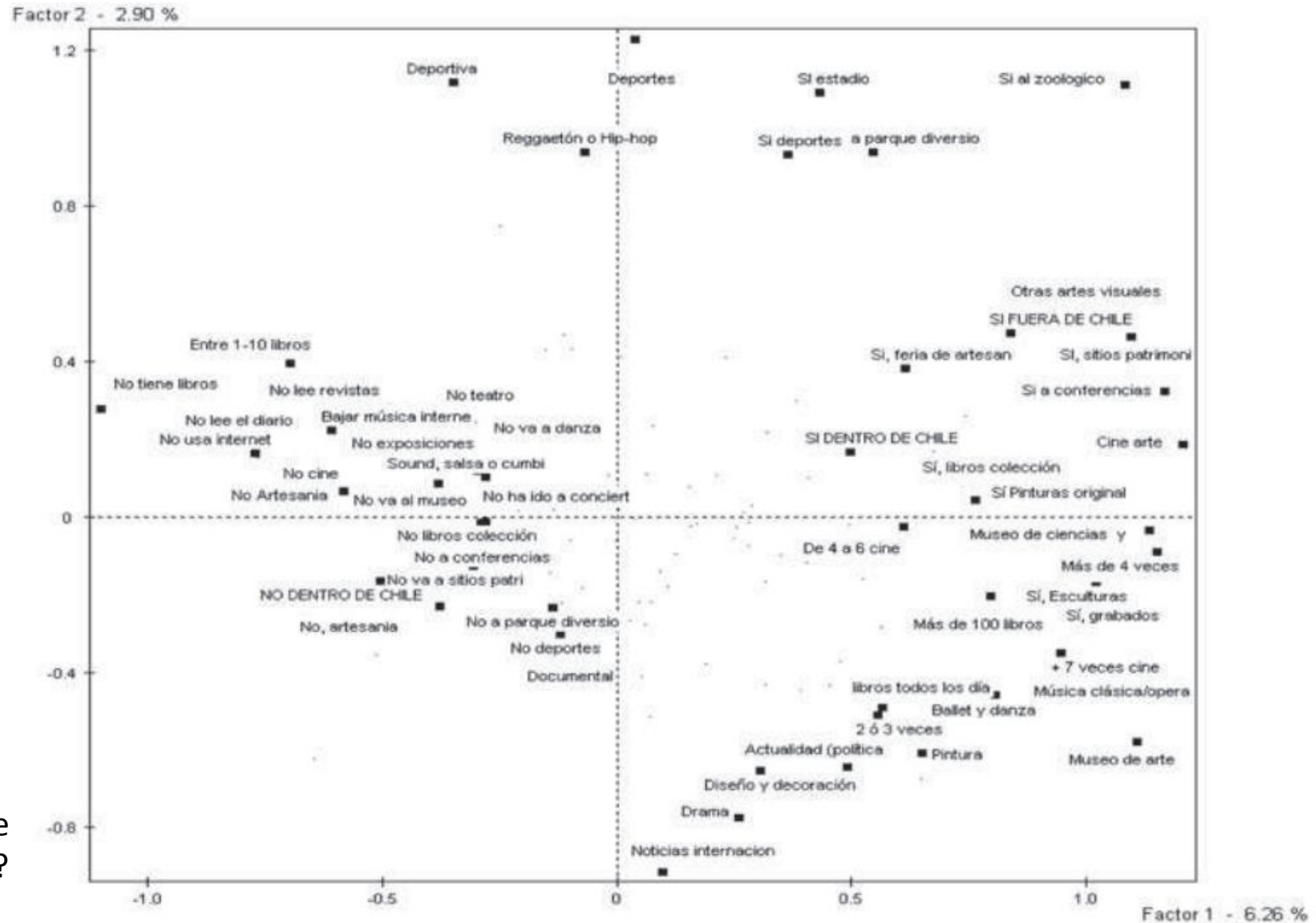
- Técnica útil para representar espacialmente, en “mapas perceptuales”, relaciones entre variables categóricas
- Como se trabaja con variables categóricas, el punto de partida de esta técnica es una tabla de contingencia
- La generación de mapas perceptuales supone que existe una relación significativa entre las variables utilizadas (chi-cuadrado)
- **Ejemplo:**
 - Gayo, M; B. Teitelboim y M. L. Méndez. 2013. Exclusividad y fragmentación: los perfiles culturales de la clase media en Chile. *Universum* 28 (1): 97-128

Análisis de correspondencias

Ejemplo:

- Gayo, M; B. Teitelboim y M. L. Méndez. 2013. Exclusividad y fragmentación: los perfiles culturales de la clase media en Chile. *Universum* 28 (1): 97-128
- Estudio *inductivo* de los “perfiles culturales” (ej. pautas de consumo cultural) que fragmentan a la clase media (pp. 103-104).
- Utilización de diversas variables categóricas:
 - Ocupación
 - Nivel educacional
 - Prácticas de recreación (ir al museo, al estadio, a conciertos, etc.)
 - Tipo de música que escucha (salsa, cumbia, etc.)
 - Cantidad de libros que posee (medido en categorías)

Gráfico 3. Patrones de participación y gusto de la clase media



I. Objetivos y propósito

- Representar gráficamente la estructura de relaciones de dos o más variables categóricas
- Identificar **pautas (patrones/perfiles)** en la relación entre las filas y las columnas (**categorías de las variables**)
- ¿Cómo?
- Generando **mapas perceptuales** que representen gráficamente dicha relación (la cual se asume que ya está chequeada)
 - Ojo: el AC **no** es una prueba de significación (para eso está Chi-cuadrado)

I. Objetivos y propósito

- Idea general del AC: “transformar lo categórico en numérico”
- *Mapas perceptuales*: se construyen asignando valores numéricos a las categorías de cada variable de modo tal de poder posicionarlas en un plano de (generalmente) dos ejes
- Dos tipos de AC
 1. Simple (dos variables)
 2. Múltiple (más de dos variables)

II. Diseño de investigación y supuestos

- En principio, el AC requiere solo de una tabla de contingencia para funcionar. Sin embargo, el/la investigador/a debe hacerse cargo de una serie de cuestiones:
 1. Interpretación de resultados (categorías filas y columnas)
 2. Definición de variables “dependientes” e “independientes”
 3. La tabla de contingencia puede incluir más de dos variables. En esos casos se lleva a cabo un AC múltiple
- En general, el AC es una técnica “libre de supuestos” (Hair et al, 2014: 527). El uso de tablas de contingencia para representar relación entre variables no métricas permite representar relaciones lineales y no lineales.

III. Construcción de los mapas perceptuales

- Esencia del AC: transformación de lo categórico en numérico:
 - Punto de partida: posicionamiento de las categorías de una variable según la distribución de sus casos en las categorías de otra variable (tabla de contingencia)
- AC: busca construir mapas perceptuales en donde las frecuencias relativas se transformen a *distancias entre puntos*.

III. Construcción de los mapas perceptuales

- “Transformar lo categórico en numérico” ¿Cómo?
- Punto de partida: posicionamiento de las categorías de una variable según la distribución de sus casos en las categorías de otra variable (tabla de contingencia)
- Ejemplo: Vivanco, 1999: Cap 5

Causas de muerte*ocupación (frecuencias)				
	Tumor	Cólera	Amebiasis	Total
Arquitecto/a	12	10	9	31
Fontanero/a	41	34	11	86
Sociólogo/a	5	3	2	10
Tasador/a	15	12	4	31
Total	72	60	25	157

III. Construcción de los mapas perceptuales

- Transformación de frecuencias absolutas a frecuencia relativas (filas o columnas)

Causas de muerte*ocupación (frecuencias)				
	Tumor	Cólera	Amebiasis	Total
Arquitecto/a	12	10	9	31
Fontanero/a	41	34	11	86
Sociólogo/a	5	3	2	10
Tasador/a	15	12	4	31
Total	72	60	25	157

Proporciones
filas

Causas de muerte*ocupación (Proporciones fila)				
	Tumor	Cólera	Amebiasis	Total
Arquitecto/a	0,39	0,32	0,29	1
Fontanero/a	0,48	0,39	0,13	1
Sociólogo/a	0,5	0,3	0,2	1
Tasador/a	0,48	0,39	0,13	1
Total	0,46	0,38	0,13	1

II. Construcción de los mapas perceptuales

- **Primer concepto clave en el AC**

- **Perfil (fila o columna):** cocientes entre frecuencias de las celdas de la variable independiente (fila, en este caso) □ base para posicionar gráficamente las puntuaciones de cada categoría

Causas de muerte*ocupación (Proporciones fila)				
	Tumor	Cólera	Amebiasis	Total
Arquitecto/a	0,39	0,32	0,29	1
Fontanero/a	0,48	0,39	0,13	1
Sociólogo/a	0,5	0,3	0,2	1
Tasador/a	0,48	0,39	0,13	1
Total	0,46	0,38	0,13	1

Ejemplos:

II. Construcción de los mapas perceptuales

- **Primer concepto clave en el AC**

- **Perfil (fila o columna):** cocientes entre frecuencias de las celdas de la variable independiente (fila, en este caso) □ base para posicionar gráficamente las puntuaciones de cada categoría

Causas de muerte*ocupación (Proporciones fila)				
	Tumor	Cólera	Amibiasis	Total
Arquitecto/a	0,39	0,32	0,29	1
Fontanero/a	0,48	0,39	0,13	1
Sociólogo/a	0,5	0,3	0,2	1
Tasador/a	0,48	0,39	0,13	1
Total	0,46	0,38	0,13	1

Ejemplos:

- **Perfil fila de Arquitecto/a:** [0,39; 0,32 y 0,29]

II. Construcción de los mapas perceptuales

- **Primer concepto clave en el AC**

- **Perfil (fila o columna):** cocientes entre frecuencias de las celdas de la variable independiente (fila, en este caso) □ base para posicionar gráficamente las puntuaciones de cada categoría

Causas de muerte*ocupación (Proporciones fila)				
	Tumor	Cólera	Amibiasis	Total
Arquitecto/a	0,39	0,32	0,29	1
Fontanero/a	0,48	0,39	0,13	1
Sociólogo/a	0,5	0,3	0,2	1
Tasador/a	0,48	0,39	0,13	1
Total	0,46	0,38	0,13	1

Ejemplos:

- **Perfil fila de Arquitecto/a:** [0,39; 0,32 y 0,29]
- **Perfil fila medio:** [0,46; 0,38 y 0,13]

II. Construcción de los mapas perceptuales

- **Primer concepto clave en el AC**

- **Perfil (fila o columna):** cocientes entre frecuencias de las celdas de la variable independiente (fila, en este caso) □ base para posicionar gráficamente las puntuaciones de cada categoría

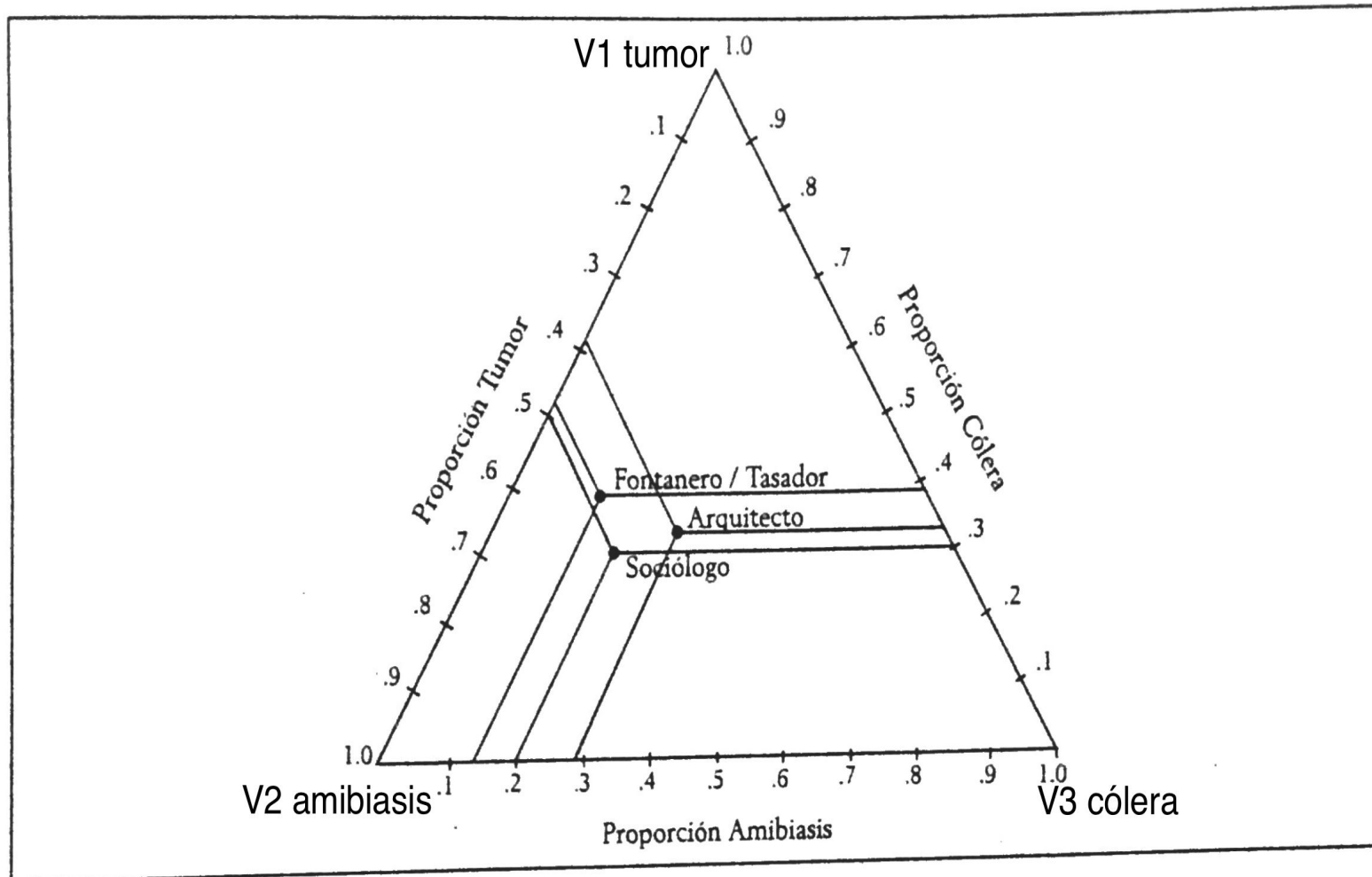
Causas de muerte*ocupación (Proporciones fila)				
	Tumor	Cólera	Amibiasis	Total
Arquitecto/a	0,39	0,32	0,29	1
Fontanero/a	0,48	0,39	0,13	1
Sociólogo/a	0,5	0,3	0,2	1
Tasador/a	0,48	0,39	0,13	1
Total	0,46	0,38	0,13	1

Ejemplos:

- **Perfil fila de Arquitecto/a:** [0,39; 0,32 y 0,29]
- **Perfil fila medio:** [0,46; 0,38 y 0,13]
- ¿Qué pasa con fontaneros/as y tasadores/as? Mismo perfil = misma posición gráfica

III. Construcción de los mapas perceptuales

- Representaciones gráficas



Vértices

- Representan las categorías de la variable dependiente (columna)
- Representan perfiles “puros”

$$V1 \text{ tumor} = [1; 0; 0]$$

$$V2 \text{ amibiasis} = [0; 1; 0]$$

$$V3 \text{ cólera} = [0; 0; 1]$$

Por lo tanto, la **posición real de un perfil va representar (por ahora) su distancia con dichos perfiles “puros”**

III. Construcción de los mapas perceptuales

- Ejemplo: Greenacre, 2008: cap. 3

Imagen 3.1:
Tabla del cruce del nivel de educación por tipo de lector, que muestra los perfiles fila y el perfil fila medio entre paréntesis, así como las masas de las filas (derivadas de los totales de las filas)

NIVEL DE EDUCACIÓN	TIPO DE LECTOR			Total	Masas de las filas
	Rápidos C1	Minuciosos C2	Muy minuciosos C3		
Educación primaria incompleta	5	7	2	14	0,045
E1	(0,357)	(0,500)	(0,143)		
Educación primaria	18	46	20	84	0,269
E2	(0,214)	(0,548)	(0,238)		
Educación secundaria incompleta	19	29	39	87	0,279
E3	(0,218)	(0,333)	(0,448)		
Educación secundaria	12	40	49	101	0,324
E4	(0,119)	(0,396)	(0,485)		
Educación universitaria incompleta	3	7	16	26	0,083
E5	(0,115)	(0,269)	(0,615)		
Total	57	129	126	312	
Perfil fila medio	(0,183)	(0,413)	(0,404)		

Imagen 3.1:
Tabla del cruce del nivel de educación por tipo de lector, que muestra los perfiles fila y el perfil fila medio entre paréntesis, así como las masas de las filas (derivadas de los totales de las filas)

NIVEL DE EDUCACIÓN	TIPO DE LECTOR			Total	Masas de las filas
	Rápidos C1	Minuciosos C2	Muy minuciosos C3		
Educación primaria incompleta E1	5 (0,357)	7 (0,500)	2 (0,143)	14	0,045
Educación primaria E2	18 (0,214)	46 (0,548)	20 (0,238)	84	0,269
Educación secundaria incompleta E3	19 (0,218)	29 (0,333)	39 (0,448)	87	0,279
Educación secundaria E4	12 (0,119)	40 (0,396)	49 (0,485)	101	0,324
Educación universitaria incompleta E5	3 (0,115)	7 (0,269)	16 (0,615)	26	0,083
Total	57	129	126	312	
Perfil fila medio	(0,183)	(0,413)	(0,404)		

2 conceptos claves:

1. Perfiles:

- cocientes entre frecuencias de las celdas de la variable independiente (fila, en este caso)
- Ej. perfil fila E1: [0,357; 0,500; 0,143]
- Perfil fila medio: [0,183; 0,413; 0,404]

Imagen 3.1:

Tabla del cruce del nivel de educación por tipo de lector, que muestra los perfiles fila y el perfil fila medio entre paréntesis, así como las masas de las filas (derivadas de los totales de las filas)

NIVEL DE EDUCACIÓN	TIPO DE LECTOR			Total	Masas de las filas
	Rápidos C1	Minuciosos C2	Muy minuciosos C3		
Educación primaria incompleta E1	5 (0,357)	7 (0,500)	2 (0,143)	14	0,045
Educación primaria E2	18 (0,214)	46 (0,548)	20 (0,238)	84	0,269
Educación secundaria incompleta E3	19 (0,218)	29 (0,333)	39 (0,448)	87	0,279
Educación secundaria E4	12 (0,119)	40 (0,396)	49 (0,485)	101	0,324
Educación universitaria incompleta E5	3 (0,115)	7 (0,269)	16 (0,615)	26	0,083
Total	57	129	126	312	
Perfil fila medio	(0,183)	(0,413)	(0,404)		

2. Vértices:

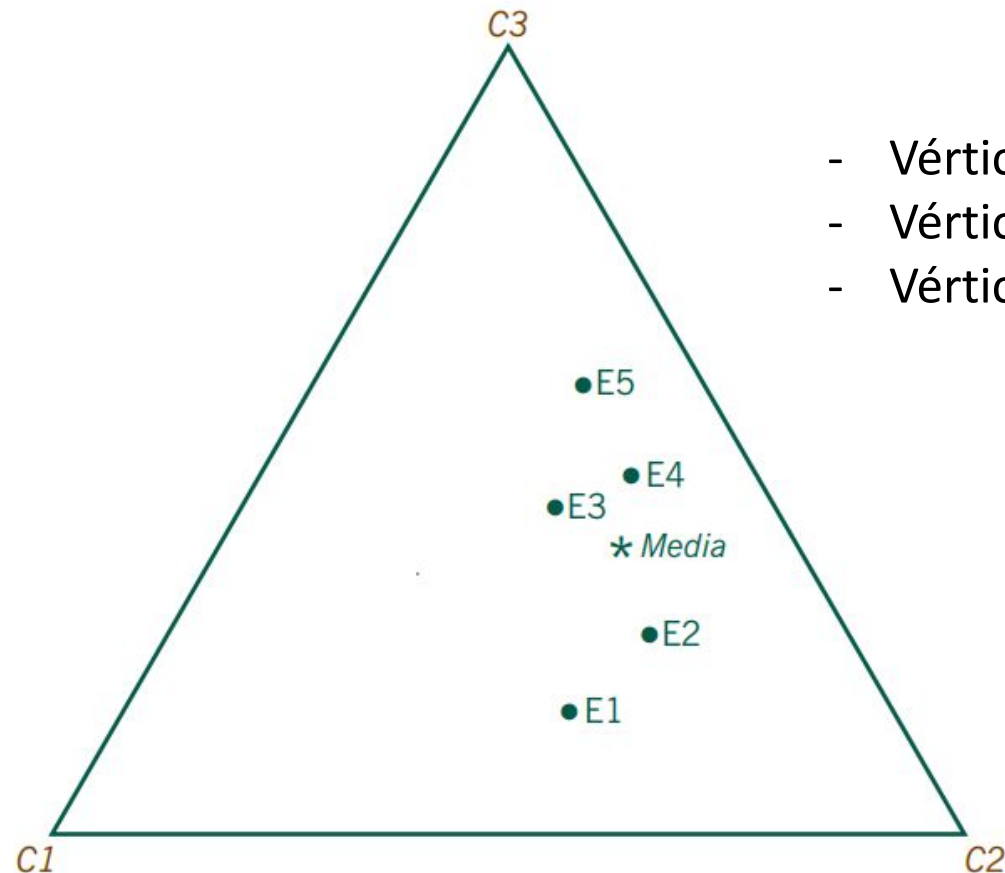
- Representan las categorías de la variable dependiente (columna) y, al mismo tiempo, perfiles “puros”
 - Vértice C1 Rápidos [1; 0; 0]
 - Vértice C2 Minuciosos [0; 1; 0]
 - Vértice C3 Muy minuciosos [0; 0; 1]

III. Construcción de los mapas perceptuales

- El análisis conjunto de perfiles y vértices es el primer paso para posicionar espacialmente las categorías de las variables analizadas

Imagen 3.2:

Representación gráfica de los perfiles fila (nivel de educación) de la imagen 3.1 en coordenadas triangulares, que también indica la posición del perfil fila medio (última fila de la imagen 3.1)

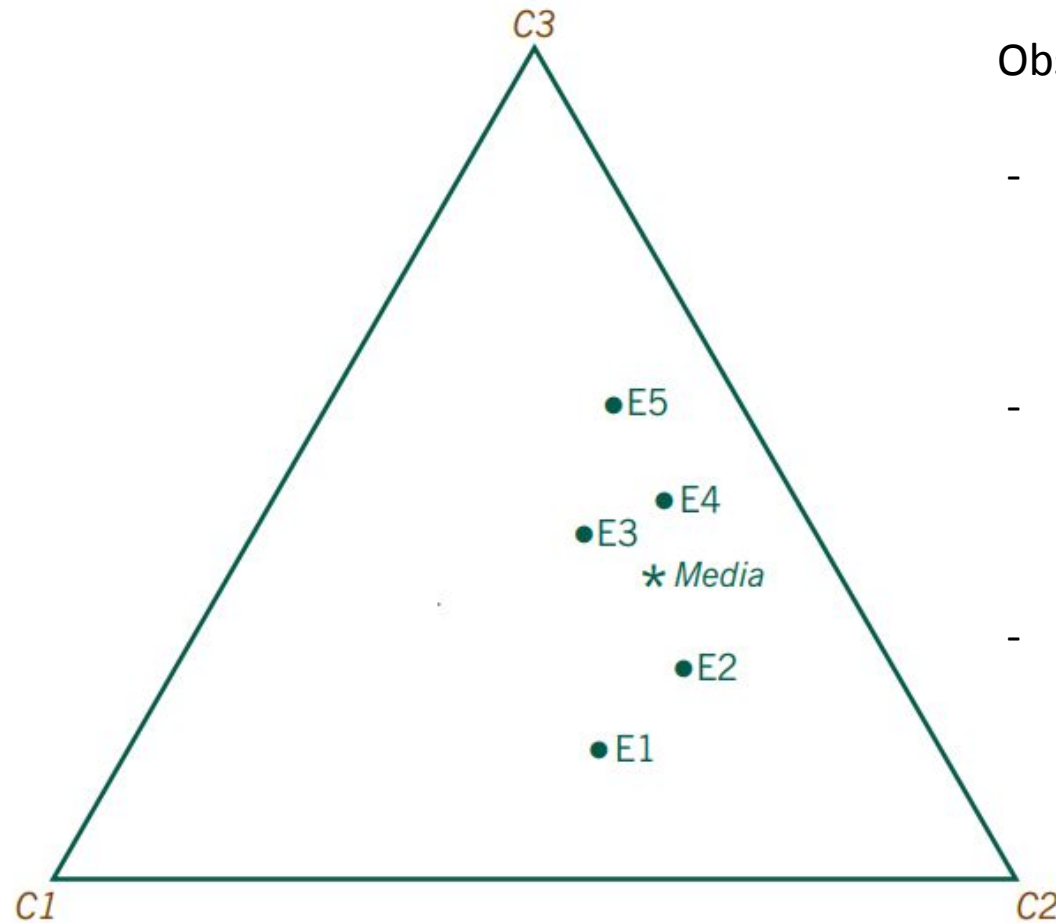


- Vértice C1 Rápidos [1; 0; 0]
- Vértice C2 Minuciosos [0; 1; 0]
- Vértice C3 Muy minuciosos [0; 0; 1]

En una representación de este tipo, la posición de un perfil depende de su distancia respecto de los vértices (perfiles “puros”)

III. Construcción de los mapas perceptuales

Imagen 3.2:
Representación gráfica de los perfiles fila (nivel de educación) de la imagen 3.1 en coordenadas triangulares, que también indica la posición del perfil fila medio (última fila de la imagen 3.1)

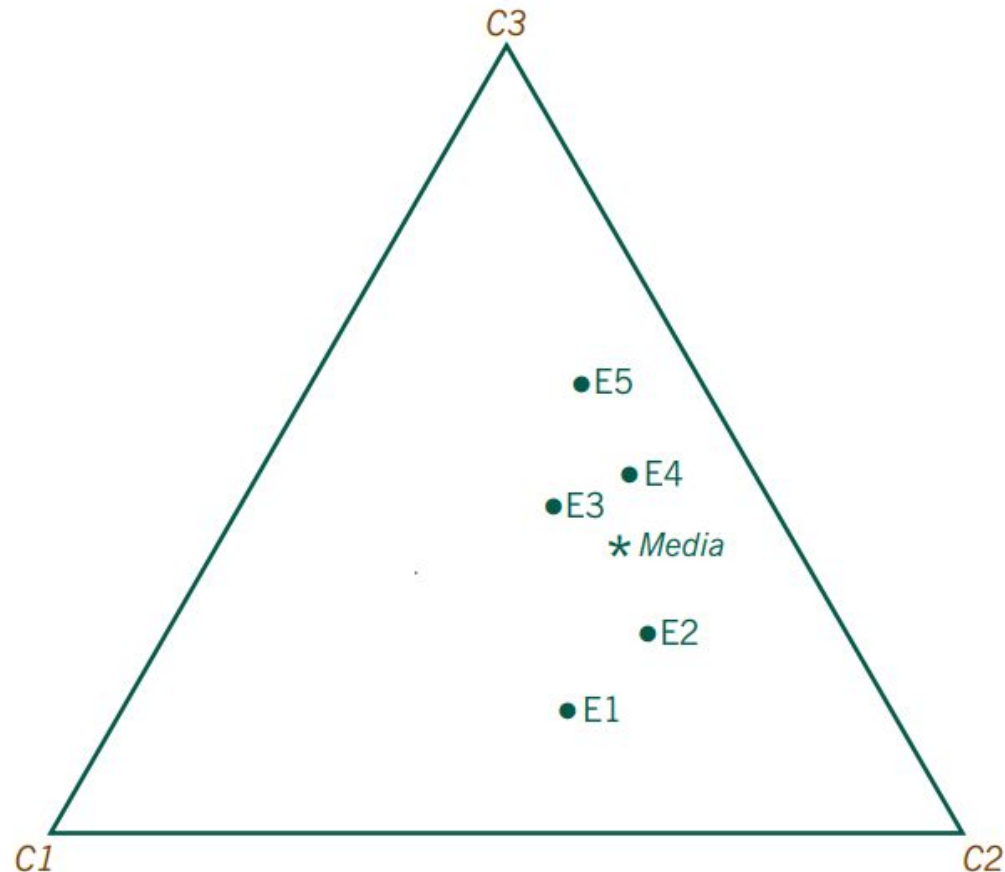


Observar en la representación gráfica:

- El **perfil E5** está más cerca del vértice C3 (más lectores “muy minuciosos” en ese grupo educacional)
- El **perfil E1** está lejos del vértice C3 (hay pocos lectores “muy minuciosos” en ese grupo ed.)
- La **media** está cargada hacia C2 y C3 (el perfil fila medio muestra que pocos casos son “lectores rápidos”)

III. Construcción de los mapas perceptuales

Imagen 3.2:
*Representación gráfica de
los perfiles fila (nivel de
educación) de la imagen 3.1
en coordenadas
triangulares, que también
indica la posición del perfil
fila medio (última fila de la
imagen 3.1)*



Pregunta clave:

- ¿Cómo sabemos que el AC genera un mapa que representa correctamente la relación entre variables?
- Respuesta: porque en el AC se trabaja con la noción de **centroide** (= **baricentro o media ponderada**)

III. Construcción de los mapas perceptuales

- Media ponderada (centroide o baricentro): punto “promedio” de cada perfil
- En el AC la media ponderada de un perfil *no* representa el promedio simple entre los porcentajes asociados a una categoría fila—por ejemplo (E5) y los valores de sus perfiles [0,115; 0,269; 0,615].
- El centroide representa el “**promedio**” entre la posición de un perfil en cada uno de los tres vértices, **ponderado** por la cantidad de casos que dicho perfil tiene en cada ***vértice***.

III. Construcción de los mapas perceptuales

- Este cálculo explica por qué la media se halla más lejos del vértice C1 (lectores rápidos): el peso de ese perfil en dicho vértice es menor que en los otros dos.

Imagen 3.2:

Representación gráfica de los perfiles fila (nivel de educación) de la imagen 3.1 en coordenadas triangulares, que también indica la posición del perfil fila medio (última fila de la imagen 3.1)

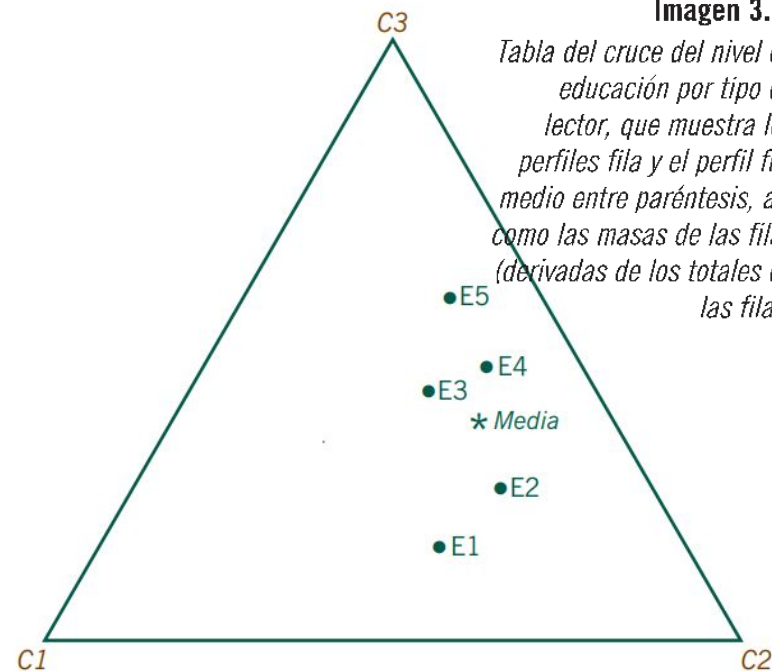


Imagen 3.1:

Tabla del cruce del nivel de educación por tipo de lector, que muestra los perfiles fila y el perfil fila medio entre paréntesis, así como las masas de las filas (derivadas de los totales de las filas)

NIVEL DE EDUCACIÓN	TIPO DE LECTOR			Total
	Rápidos C1	Minuciosos C2	Muy minuciosos C3	
Educación primaria incompleta E1	5 (0,357)	7 (0,500)	2 (0,143)	14
Educación primaria E2	18 (0,214)	46 (0,548)	20 (0,238)	84
Educación secundaria incompleta E3	19 (0,218)	29 (0,333)	39 (0,448)	87
Educación secundaria E4	12 (0,119)	40 (0,396)	49 (0,485)	101
Educación universitaria incompleta E5	3 (0,115)	7 (0,269)	16 (0,615)	26
Total	57	129	126	312
Perfil fila medio	(0,183)	(0,413)	(0,404)	

¿En qué circunstancias la media (*) se ubicaría justo al centro del triángulo? Cuando el perfil medio sea [0,333; 0,333; 0,333]

III. Construcción de los mapas perceptuales

Perfil medio como “caso especial” en el AC

- En el AC el perfil medio no es sólo la media ponderada (centroide) de los tres vértices
- **El perfil medio es también el centroide de los cinco perfiles fila** (la *media ponderada* de las 5 categorías de educación)
- En este caso, la posición del perfil fila medio puede ser calculada a partir de *las masas* asociadas a cada perfil
 - Otro concepto importante: **masa**

III. Construcción de los mapas perceptuales

Imagen 3.1:

Tabla del cruce del nivel de educación por tipo de lector, que muestra los perfiles fila y el perfil fila medio entre paréntesis, así como las masas de las filas (derivadas de los totales de las filas)

NIVEL DE EDUCACIÓN	TIPO DE LECTOR			Total	<u>Masas de las filas</u>
	Rápidos C1	Minuciosos C2	Muy minuciosos C3		
Educación primaria incompleta E1	5 (0,357)	7 (0,500)	2 (0,143)	14	0,045
Educación primaria E2	18 (0,214)	46 (0,548)	20 (0,238)	84	0,269
Educación secundaria incompleta E3	19 (0,218)	29 (0,333)	39 (0,448)	87	0,279
Educación secundaria E4	12 (0,119)	40 (0,396)	49 (0,485)	101	0,324
Educación universitaria incompleta E5	3 (0,115)	7 (0,269)	16 (0,615)	26	0,083
Total	57	129	126	312	
Perfil fila medio	(0,183)	(0,413)	(0,404)		

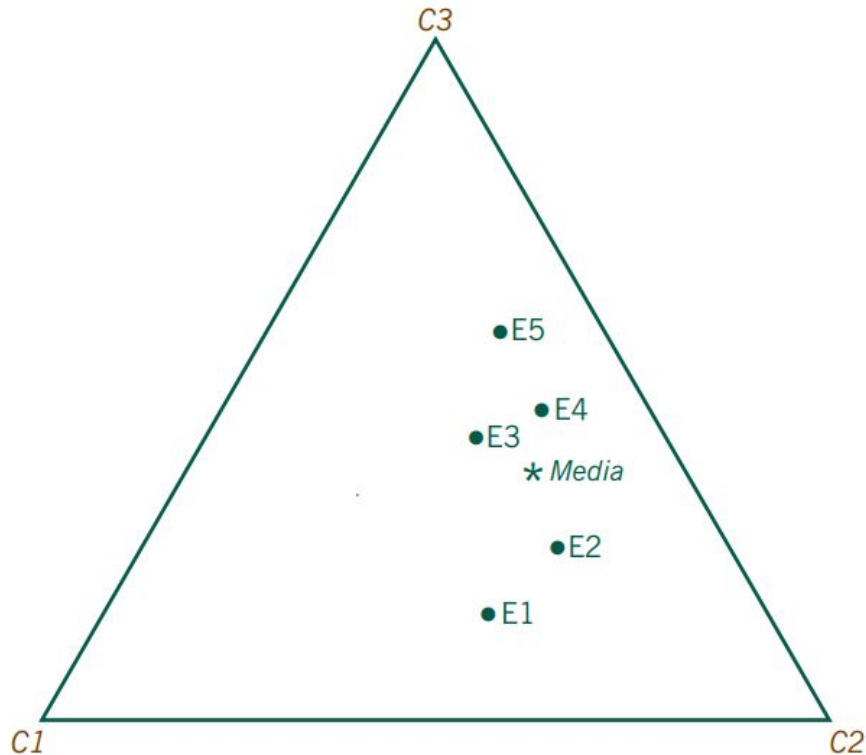
Concepto de masa

- Representa el “peso” que tiene cada fila en relación con el total de casos (frecuencias marginales de cada fila / 312)
- A partir de las masas se puede corregir el cálculo de la posición del perfil fila medio (de la suma “hacia abajo”) considerando el hecho que cada perfil fila tiene distintos tamaños
 - E1 = 14
 - E2 = 84
 - E3 = 87
 - Etc.

III. Construcción de los mapas perceptuales

- A partir de **las masas** se **obtiene la posición del perfil fila medio**.
- Esta se halla al “centro” de los perfiles fila, pero **más cerca de los perfiles con mayor frecuencia** (ej. perfil E3).

Imagen 3.2:
*Representación gráfica de
los perfiles fila (nivel de
educación) de la imagen 3.1
en coordenadas
triangulares, que también
indica la posición del perfil
fila medio (última fila de la
imagen 3.1)*



IV. Distancia Chi-cuadrado e inercia

- En el AC las distancias entre los perfiles no se miden directamente (como en el ejemplo anterior), sino que a través de la distancia Chi-cuadrado.
- Recordar Chi-cuadrado:

Imagen 4.1:
Frecuencias observadas, tal como aparecen en la imagen 3.1 junto con las frecuencias esperadas (entre paréntesis) calculadas suponiendo que se cumple el supuesto de homogeneidad

NIVEL DE EDUCACIÓN	TIPO DE LECTOR			Total	Masas de las filas
	Rápidos C1	Minuciosos C2	Muy minuciosos C3		
Educación primaria incompleta E1	5 (2,56)	7 (5,78)	2 (5,66)	14	0,045
Educación primaria E2	18 (15,37)	46 (34,69)	20 (33,94)	84	0,269
Educación secundaria incompleta E3	19 (15,92)	29 (35,93)	39 (35,15)	87	0,279
Educación secundaria E4	12 (18,48)	40 (41,71)	49 (40,80)	101	0,324
Educación universitaria incompleta E5	3 (4,76)	7 (10,74)	16 (10,50)	26	0,083
Total	57	129	126	312	
Perfil fila medio	(0,183)	(0,413)	(0,404)		

IV. Distancia Chi-cuadrado e inercia

- En el AC, el Chi-cuadrado *no* interesa como prueba de significancia, sino como **medida de la heterogeneidad entre los perfiles** (es decir, como medida de su “distancia” respecto del perfil fila medio)
- Para ello, hay que hacer 2 transformaciones del estadístico χ^2 (Greenacre, 2008: 47-48)

IV. Distancia Chi-cuadrado e inercia

- Transformaciones del estadístico χ^2

1. Cada término de la fórmula de χ^2 se divide por el cuadrado del total de cada fila ☐

Resultado: ya no se comparan F_o y F_e , sino perfiles observados con perfiles esperados.

- El perfil fila medio funciona ahora como “perfil esperado”

Matemáticamente, esto significa que:

$$\chi^2 \text{ cada fila} = \text{total de la fila} * \frac{(\text{perfiles fila observados} - \text{perfiles fila esperados})^2}{\text{perfiles fila esperados}}$$

IV. Distancia Chi-cuadrado e inercia

1. Ejemplo, celdas de la fila E5

Imagen 3.1:

Tabla del cruce del nivel de educación por tipo de lector, que muestra los perfiles fila y el perfil fila medio entre paréntesis, así como las masas de las filas (derivadas de los totales de las filas)

NIVEL DE EDUCACIÓN	TIPO DE LECTOR			Total	Masas de las filas
	Rápidos C1	Minuciosos C2	Muy minuciosos C3		
Educación primaria incompleta E1	5 (0,357)	7 (0,500)	2 (0,143)	14	0,045
Educación primaria E2	18 (0,214)	46 (0,548)	20 (0,238)	84	0,269
Educación secundaria incompleta E3	19 (0,218)	29 (0,333)	39 (0,448)	87	0,279
Educación secundaria E4	12 (0,119)	40 (0,396)	49 (0,485)	101	0,324
Educación universitaria incompleta E5	3 (0,115)	7 (0,269)	16 (0,615)	26	0,083
Total	57	129	126	312	
Perfil fila medio	(0,183)	(0,413)	(0,404)		

$$E5 = 26 * \frac{(0,115 - 0,183)^2}{0,183} + 26 * \frac{(0,269 - 0,413)^2}{0,413} + 26 * \frac{(0,615 - 0,404)^2}{0,404}$$

Donde:

- 0,115 = perfil fila observado celda E5-C1 = $F_o / \text{total fila} = 3/26$
- 0,183 = perfil fila medio celda E5-C1 =
 - Esto es lo mismo que $F_e / \text{total fila} = 4,76/26 = 0,183$ (lámina siguiente)

En suma: la división de cada componente del perfil medio con el total **es lo mismo** que la división de cada componente de cada perfil esperado (E1, E2, etc.) con su total respectivo

IV. Distancia Chi-cuadrado e inercia

1. Ejemplo, celdas de la fila E5

Imagen 4.1:
Frecuencias observadas, tal como aparecen en la imagen 3.1 junto con las frecuencias esperadas (entre paréntesis) calculadas suponiendo que se cumple el supuesto de homogeneidad

NIVEL DE EDUCACIÓN	TIPO DE LECTOR			Total	Masas de las filas
	Rápidos C1	Minuciosos C2	Muy minuciosos C3		
Educación primaria incompleta E1	5 (2,56)	7 (5,78)	2 (5,66)	14	0,045
Educación primaria E2	18 (15,37)	46 (34,69)	20 (33,94)	84	0,269
Educación secundaria incompleta E3	19 (15,92)	29 (35,93)	39 (35,15)	87	0,279
Educación secundaria E4	12 (18,48)	40 (41,71)	49 (40,80)	101	0,324
Educación universitaria incompleta E5	3 (4,76)	7 (10,74)	16 (10,50)	26	0,083
Total	57	129	126	312	
Perfil fila medio	(0,183)	(0,413)	(0,404)		

$$Fe_{E5-C1} / \text{total fila} = 4,76/26 = 0,183$$

IV. Distancia Chi-cuadrado e inercia

- Transformaciones del estadístico χ^2

1. Comparación entre perfiles observados y esperados

Consecuencia para Chi-cuadrado: la fórmula queda con un término “excedente” (total de filas que multiplica la resta al cuadrado)

$$\chi^2 = \sum \text{total filas} * \frac{(\text{perfiles fila observados} - \text{perfiles fila esperados})^2}{\text{perfiles fila esperados}}$$

Esto hace necesario una segunda transformación del estadístico χ^2

IV. Distancia Chi-cuadrado e inercia

2. Corrección de este término excedente: **se dividen los dos lados de la ecuación por el total de la muestra.**

$$\frac{\chi^2}{n} = \sum \frac{\text{total filas}}{n} * \frac{(\text{perfiles fila observados} - \text{perfiles fila esperados})^2}{\text{perfiles fila esperados}}$$

- ¿Qué se logra con esto?
 - Que la parte derecha de **la ecuación** represente un factor que no se ve afectado por el total de la fila sino que está **ponderado por su masa** (es decir, el peso que tiene cada fila en relación al total).

IV. Distancia Chi-cuadrado e inercia

- Transformaciones del estadístico χ^2

2. Ejemplo fila E5:

$$E5 = \frac{26}{312} * \frac{(0,115 - 0,183)^2}{0,183} + \frac{26}{312} * \frac{(0,269 - 0,413)^2}{0,413} + \frac{26}{312} * \frac{(0,615 - 0,404)^2}{0,404}$$

$$E5 = 0,083 * \left[\frac{(0,115 - 0,183)^2}{0,183} + \frac{(0,269 - 0,413)^2}{0,413} + \frac{(0,615 - 0,404)^2}{0,404} \right]$$

Notar que 0,083 = masa perfil E5 (ver lámina siguiente)

IV. Distancia Chi-cuadrado e inercia

- Transformaciones del estadístico χ^2

LO REALMENTE IMPORTANTE

- Esta segunda transformación se hace para medir la “contribución” de cada perfil a la varianza total de la tabla.
- Esta idea es la base del concepto de inercia, uno los **conceptos más importantes del AC**

IV. Distancia Chi-cuadrado e inercia

- Transformaciones del estadístico χ^2

Imagen 3.1:

Tabla del cruce del nivel de educación por tipo de lector, que muestra los perfiles fila y el perfil fila medio entre paréntesis, así como las masas de las filas (derivadas de los totales de las filas)

NIVEL DE EDUCACIÓN	TIPO DE LECTOR			Total	<u>Masas de las filas</u>
	Rápidos C1	Minuciosos C2	Muy minuciosos C3		
Educación primaria incompleta E1	5 (0,357)	7 (0,500)	2 (0,143)	14	0,045
Educación primaria E2	18 (0,214)	46 (0,548)	20 (0,238)	84	0,269
Educación secundaria incompleta E3	19 (0,218)	29 (0,333)	39 (0,448)	87	0,279
Educación secundaria E4	12 (0,119)	40 (0,396)	49 (0,485)	101	0,324
Educación universitaria incompleta E5	3 (0,115)	7 (0,269)	16 (0,615)	26	0,083
Total	57	129	126	312	
Perfil fila medio	(0,183)	(0,413)	(0,404)		

IV. Distancia Chi-cuadrado e inercia

- **Inercia**
- Corresponde al valor χ^2/n y representa, por tanto, una **medida de la varianza total de la tabla independiente de su tamaño**
- **Varianza** = distancia total entre los perfiles y su perfil fila medio o sus perfiles esperados.

$$inercia = \frac{\chi^2}{n}$$

- En estadística, este valor también recibe el nombre de “coeficiente medio cuadrático de contingencia”

IV. Distancia Chi-cuadrado e inercia

- **Inercia y distancias**
- En el AC las distancias Chi-cuadrado entre perfiles fila y el perfil fila medio se ubican en un plano definido por los vértices de la variable columna de modo similar a como se hace cuando se trabaja con distancias euclídeas
- Sin embargo, acá hay un elemento extra no considerado en la distancia euclídea: cada distancia está dividida por los elementos esperados de cada perfil .
- Así dicha distancia es ponderada según la contribución realizada por cada celda a la suma de las distancias totales.

IV. Distancia Chi-cuadrado e inercia

Inercia y masas

- En el en el AC dichas distancias χ^2 están corregidas, además, por su masa respectiva. Al hacer eso, entonces se tiene que una nueva forma de expresar la inercia:

$$Inercia = \sum masas_{filas} * Distancia \chi^2_{filas}$$

- **Suma de las masas = 1.** Entonces, se puede entender la inercia como la media ponderada de las distancias χ^2 entre los perfiles fila y su perfil fila medio
- El **valor de la inercia** es igual a la dimensionalidad del espacio (ej. en un espacio de dos dimensiones la inercia será igual a 2)