

# **Machine Learning aplicado a las Ciencias Sociales**

## **Clase 1. Introducción y análisis no supervisado 2 (MCA)**



# Tipos de problemas

- **Aprendizaje Supervisado**
  - Variable dependiente  $\Rightarrow Y$ , resultado, target
  - Matriz de predictores ( $p$ ),  $X$ , features, variables independientes, etc.
  - Problemas de regresión:  $Y$  es cuantitativa
  - Problemas de regresión:  $Y$  es cualitativa
  - Tenemos datos de entrenamiento (conjuntos de  $X_i, Y_i$ ), observaciones
  - Podemos definir una (o varias) métricas para evaluar los modelos
- **Aprendizaje no Supervisado**
  - **No hay variable target ( $Y$ )**
  - **Solo hay  $X$**
  - **Es más difícil evaluar qué tan bien funciona el modelo**

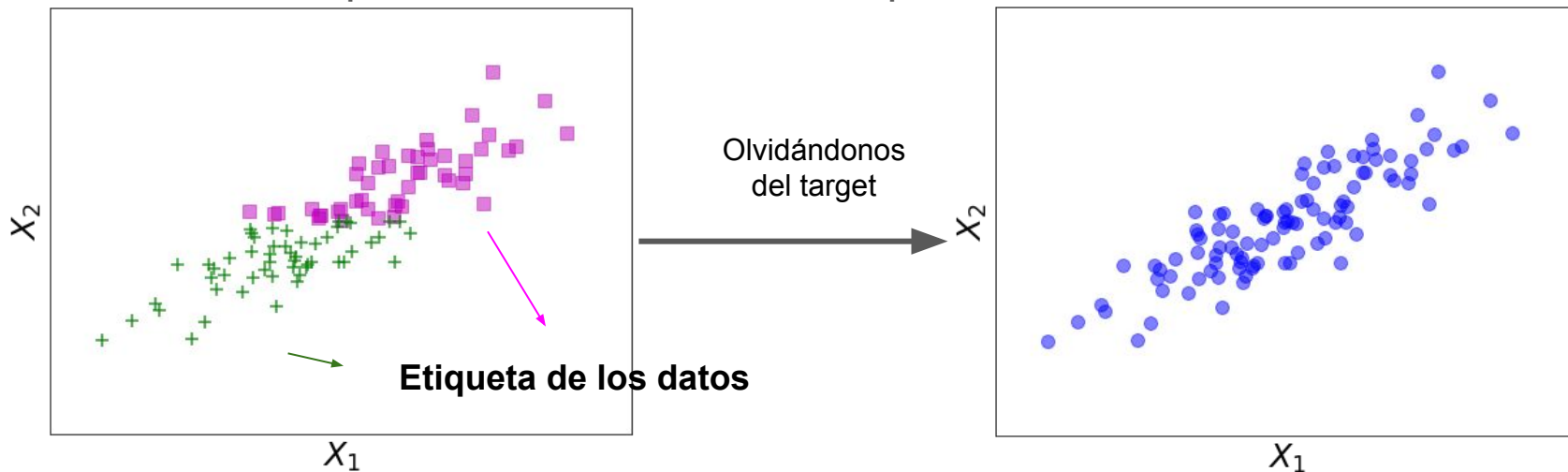


# Aprendizaje No Supervisado



# ¿Qué es el aprendizaje no supervisado?

- El estudio o la exploración de cómo están representados datos y qué conclusiones podemos sacar de dicha representación.



**Aprendizaje no supervisado** es todo lo que podemos hacer con solo la representación de los datos en el espacio de features.



# Análisis de Correspondencias Múltiples (ACM)

Generalización de AC

Individuos	Género	Años	Ingreso
1	Mujer	5	Medio
2	Mujer	3	Alto
3	Hombre	4	Bajo
4	Mujer	1	Bajo
5	Mujer	2	Medio
6	Hombre	5	Alto
7	Mujer	2	Medio
8	Hombre	3	Bajo
9	Hombre	1	Alto
10	Mujer	4	Medio



# Análisis de Correspondencias Múltiples (ACM)

## Generalización de AC

A partir de la tabla original se construye la *tabla disyuntiva (matriz Z)* con tantas columnas como categorías:

Género		Años					Ingresos		
Mujer	Hombre	1	2	3	4	5	Bajo	Medio	Alto
1	0	0	0	0	0	1	0	1	0
1	0	0	0	1	0	0	0	0	1
0	1	0	0	0	1	0	1	0	0
1	0	1	0	0	0	0	1	0	0
1	0	0	1	0	0	0	0	1	0
0	1	0	0	0	0	1	0	0	1
1	0	0	1	0	0	0	0	1	0
0	1	0	0	1	0	0	1	0	0
0	1	1	0	0	0	0	0	0	1
1	0	0	0	0	1	0	0	1	0

En la tabla disyuntiva completa (matriz Z), si hay alguna variable continua, debe transformarse en nominal, ordenándose en intervalos a los que se da un rango de valores.

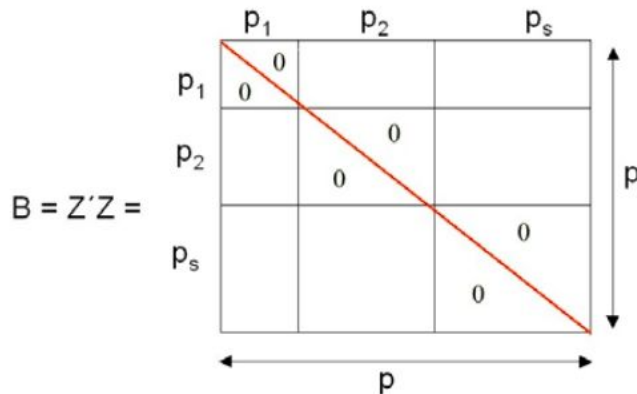


# Análisis de Correspondencias Múltiples (ACM)

## Generalización de AC

A partir de la tabla disyuntiva completa se puede construir la tabla de contingencia de Burt (B), que es una tabla simétrica de orden  $(p, p)$ :  $B = Z' \cdot Z$

B es una yuxtaposición de tablas de contingencia y está formada de  $s^2$  bloques de la forma:



Cada bloque es una submatriz formada por tablas de contingencia de las variables dos a dos, salvo los bloques que se están en la diagonal que son las tablas de contingencia de cada variable consigo misma.



# Análisis de Correspondencias Múltiples (ACM)

Tabla disyuntiva completa  
y Burt producen los  
mismos factores.

MATRIZ DE BURT

		Género		Años					Ingresos		
		M	H	1	2	3	4	5	B	M	A
Género	M	6	0	1	2	1	1	1	1	4	1
	H	0	4	1	0	1	1	1	2	0	2
Años	1	1	1	2	0	0	0	0	1	0	1
	2	2	0	0	2	0	0	0	0	2	0
	3	1	1	0	0	2	0	0	1	0	1
	4	1	1	0	0	0	2	0	1	1	0
	5	1	1	0	0	0	0	2	0	1	1
Ingresos	B	1	2	1	0	1	1	0	3	0	0
	M	4	0	0	2	0	1	1	0	4	0
	A	1	2	1	0	1	0	1	0	0	3





# Análisis de Correspondencias Múltiples (ACM)

Se puede diagonalizar la tabla de Burt para obtener los factores

MATRIZ DE BURT

		Género		Años					Ingresos		
		M	H	1	2	3	4	5	B	M	A
Género	M	6	0	1	2	1	1	1	1	4	1
	H	0	4	1	0	1	1	1	2	0	2
Años	1	1	1	2	0	0	0	0	1	0	1
	2	2	0	0	2	0	0	0	0	2	0
	3	1	1	0	0	2	0	0	1	0	1
	4	1	1	0	0	0	2	0	1	1	0
	5	1	1	0	0	0	0	2	0	1	1
Ingresos	B	1	2	1	0	1	1	0	3	0	0
	M	4	0	0	2	0	1	1	0	4	0
	A	1	2	1	0	1	0	1	0	0	3

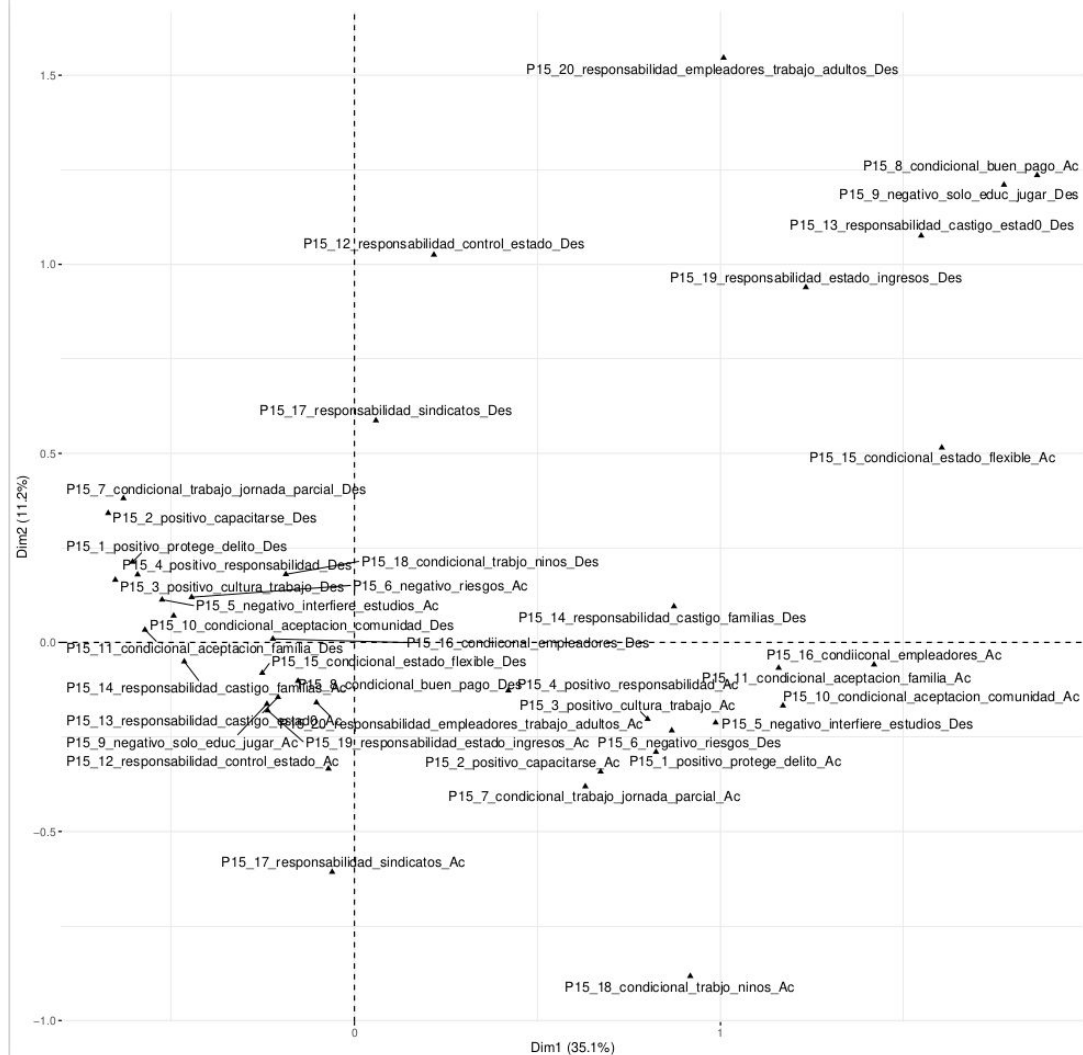


15. Por favor, dígame si Ud. está muy en desacuerdo (1), en desacuerdo (2), de acuerdo (3) o muy de acuerdo (4), con las siguientes afirmaciones...(LEER CADA ÍTEM, MOSTRAR LA TARJETA CORRESPONDIENTE Y MARCAR UNA RESPUESTA PARA CADA UNO)

AFIRMACIONES	MD	D	A	MA
<i>Es bueno que los niños pobres trabajen porque si no se vuelcan al delito y/o a las drogas</i>	1	2	3	4
<i>Es bueno que los niños trabajen porque les permite capacitarse en diferentes oficios y habilidades</i>	1	2	3	4
<i>Es bueno que las personas trabajen desde pequeñas porque se forman en la cultura del trabajo</i>	1	2	3	4
<i>Puede ser deseable que un niño combine la escuela con algún trabajo, así aprende a ser responsable</i>	1	2	3	4
<i>Los niños no deben trabajar porque interfiere con sus estudios</i>	1	2	3	4
<i>Los niños no deben trabajar porque se exponen a muchos riesgos</i>	1	2	3	4
<i>Que un niño trabaje un par de horas por día está bien pero no deberían trabajar una jornada completa</i>	1	2	3	4
<i>Si se les paga bien, debería permitirse trabajar a los niños</i>	1	2	3	4
<i>Los niños sólo deben educarse y jugar.</i>	1	2	3	4
<i>Si en la comunidad se acostumbra a que los niños trabajen está bien que lo hagan.</i>	1	2	3	4
<i>Si la familia considera que está bien que un niño trabaje, entonces, es correcto.</i>	1	2	3	4
<i>El Estado debería generar mecanismos de control adecuados para regular el trabajo infantil.</i>	1	2	3	4
<i>El Estado debería castigar a los que contratan niños para trabajar</i>	1	2	3	4
<i>Debería castigarse a las familias que hacen a sus hijos trabajar.</i>	1	2	3	4
<i>El Estado debería facilitar a los empleadores la contratación de niños para que trabajen.</i>	1	2	3	4
<i>Si los empleadores consideran que para una tarea es necesario contratar niños, entonces debería permitirselo.</i>	1	2	3	4
<i>Los sindicatos deberían intervenir para regular el trabajo infantil.</i>	1	2	3	4
<i>Existen ciertos trabajos que solamente pueden ser realizados de forma correcta por niños.</i>	1	2	3	4
<i>El Estado debe generar políticas de apoyo a las familias de menores recursos para que los niños no tengan que trabajar</i>	1	2	3	4
<i>Los empleadores deberían mejorar las condiciones de trabajo para los adultos para que los niños no tengan que trabajar</i>	1	2	3	4



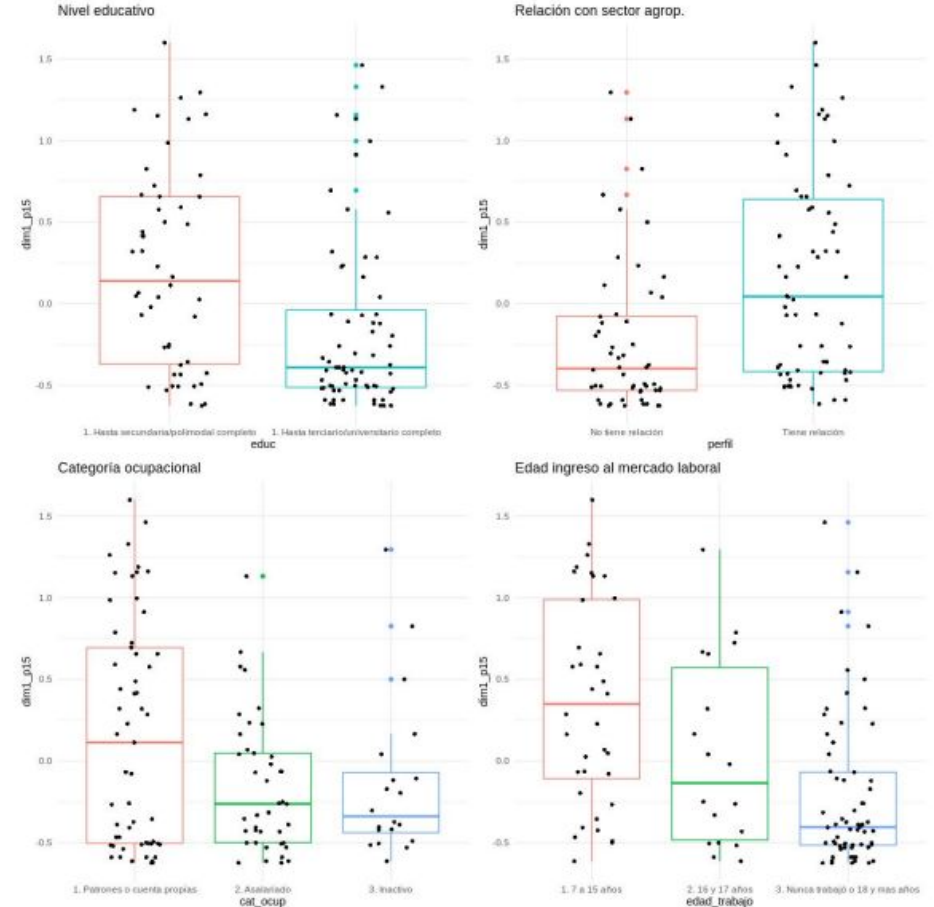
- Las categorías cercanas entre sí, implican perfiles de respuestas similares.
- Cuanto más alejadas del origen se encuentren las categorías mejor “representadas” se encuentran en la dimensión.

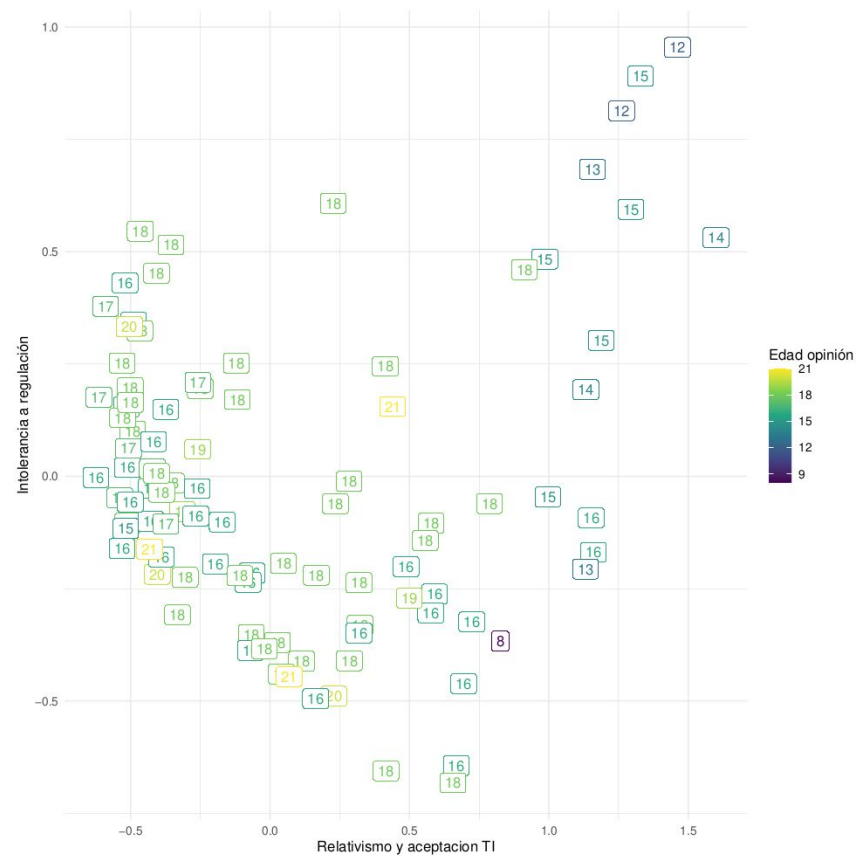
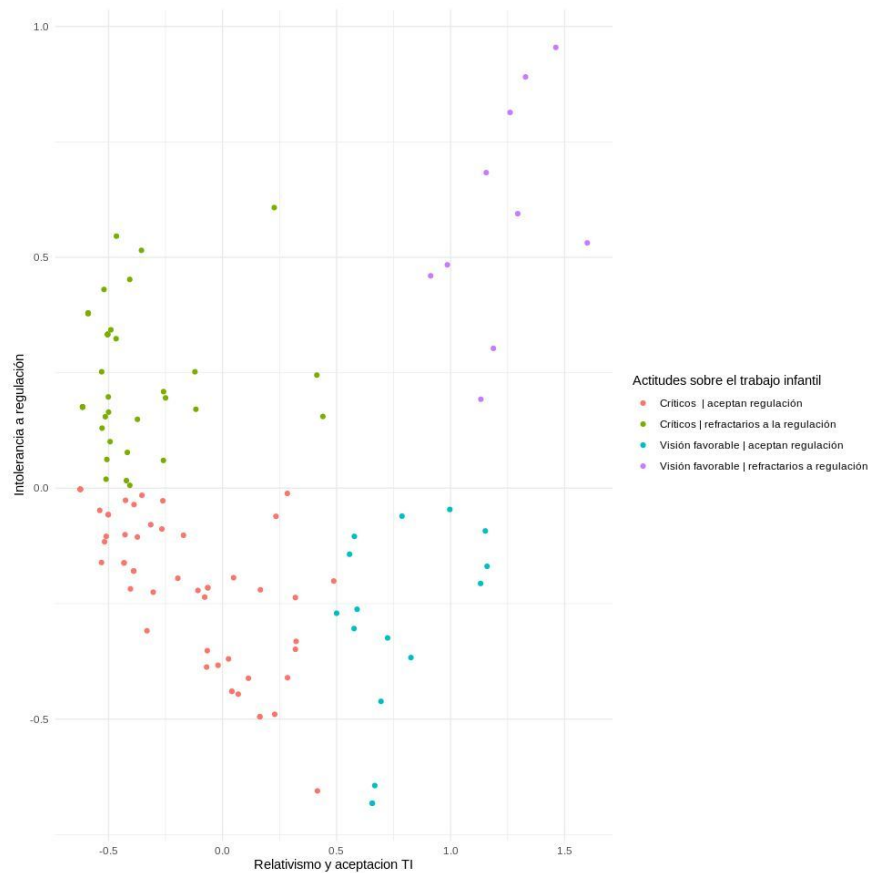


**Gráfico 12. Boxplots del grado de relativismo y acuerdo respecto al trabajo infantil según variables sociodemográficas básicas**

Los perfiles con mayor propensión a tener niveles más altos de relativismo y acuerdo con el trabajo infantil son:

- Personas con niveles de instrucción formal más bajos
- Población que desarrolla actividades laborales en el sector agropecuario
- Patrones y cuentapropistas
- Ingresantes tempranos al mercado de trabajo





# Resumen de MCA

- Análogo a PCA para variables ordinales.
- Se computan a partir de la Tabla Disyuntiva Completa o la Matriz de Burt
- Las componentes están ordenadas de mayor a menor, en el sentido de la información (inercia) que se llevan. Si tiramos las últimas componentes, estamos reduciendo la dimensión de nuestro problema.



# Vamos al Notebook



factor-data  
EIDAES\_UNSAM