

Análisis Multivariado de Datos

Estadística Inferencial Multivariada

Según la definición clásica: 2 o más variables dependientes, no importa el número de variables independientes.

Según la definición práctica: 2 o más variables independientes y 1 o más variables dependientes.

Análisis Multivariado de Datos

Análisis de varianza

- Análisis factorial de varianza
- Análisis de varianza de medidas repetidas
- Análisis de covarianza
- Análisis de varianza multivariado

Métodos de clasificación

- Análisis de regresión
- Análisis discriminante
- Análisis de regresión logística

Métodos de agrupamiento

- Análisis factorial exploratorio
- Análisis de cúmulos (clusters)

Modelos estructurales de ecuaciones

- Análisis de trayectorias (Path analysis)
- FASEM completo

Análisis de varianza

Análisis factorial de varianza

	VI_1	
	\bar{x}_1	\bar{x}_2
VI_2	\bar{x}_3	\bar{x}_4
	\bar{x}_5	\bar{x}_6

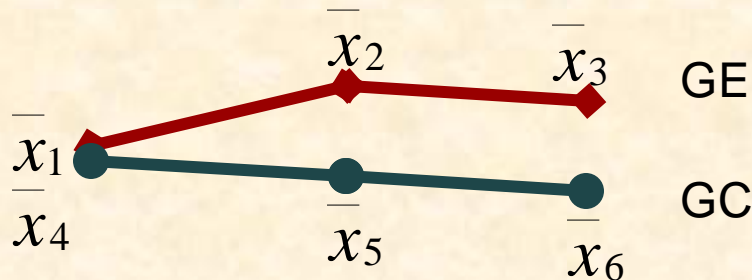
- Existe más de una variable categórica independiente o de agrupación
- Una variable de medición o dependiente medida a nivel intervalar o de razón
- En general debe cumplir los requisitos de la estadística paramétrica

Diseño 2 x 3

Por ejemplo: se evalúa un tratamiento (VI_1) para el manejo del estrés (VD), se asignan pacientes al azar al grupo control y al grupo experimental, además el nivel socioeconómico (VI_2), clasificado como bajo, medio y alto, influye sobre el resultado del tratamiento.

Análisis de varianza

Análisis de varianza de medidas repetidas



Diseño con grupo control y pretest-posttest

Por ejemplo: se evalúa un tratamiento (VI_1) para el manejo del estrés, se asignan pacientes al azar al grupo control y al grupo experimental, se evalúa el nivel de estrés (VD) antes y después del tratamiento, además una medición de seguimiento seis meses después de la finalización (VI_2).

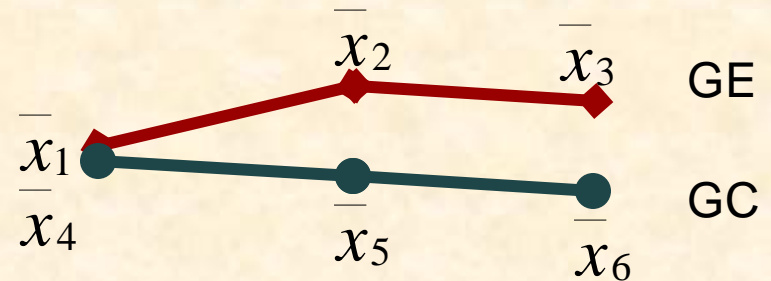
- La variable independiente o de agrupación involucra tres o más mediciones
- Puede existir una, o más de una, variable categórica de agrupación
- Una variable de medición o dependiente medida a nivel intervalar o de razón
- En general debe cumplir los requisitos de la estadística paramétrica

Análisis de varianza

Análisis de covarianza

	VI_1	
	\bar{x}_1	\bar{x}_2
VI_2	\bar{x}_3	\bar{x}_4
	\bar{x}_5	\bar{x}_6

Una/o más Variable/s
continua/s que influyen
sobre la VD



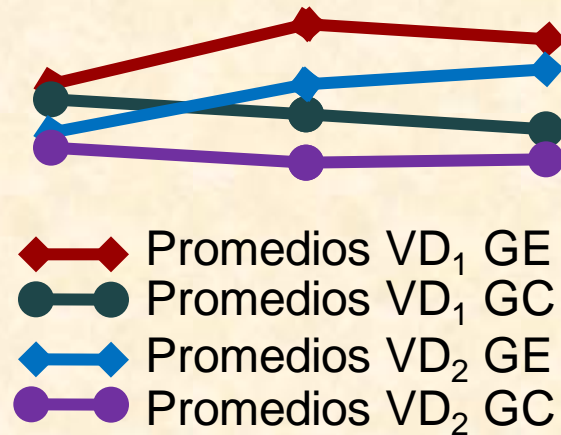
Una/o más Variable/s
continua/s que influyen
sobre la VD

Cuando en el diseño factorial de varianza o en el de medidas repetidas, se incluyen una o más variables continuas, estas se incluyen en el modelo como covariables.

Análisis de varianza

Análisis de varianza multivariado

	VI_1			
VI_2	$\bar{x}_{1_{VD1}}$	$\bar{x}_{1_{VD2}}$	$\bar{x}_{2_{VD1}}$	$\bar{x}_{2_{VD2}}$
	$\bar{x}_{3_{VD1}}$	$\bar{x}_{3_{VD2}}$	$\bar{x}_{4_{VD1}}$	$\bar{x}_{4_{VD2}}$
	$\bar{x}_{5_{VD1}}$	$\bar{x}_{5_{VD2}}$	$\bar{x}_{6_{VD1}}$	$\bar{x}_{6_{VD2}}$



Cuando en el diseño factorial de varianza o en el de medidas repetidas, se incluyen más de una variables dependientes se tiene un modelo multivariado que cumple con la definición clásica de estadística multivariada.

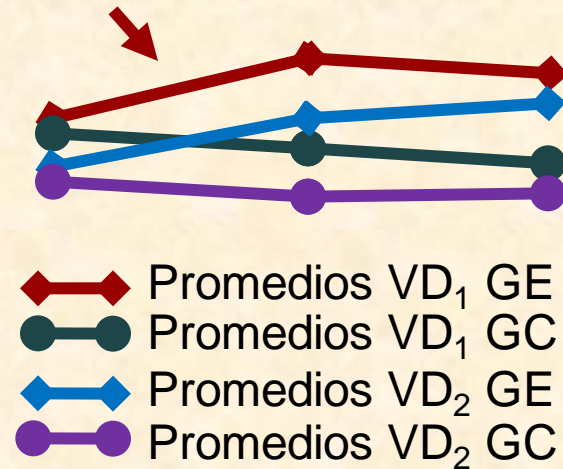
Análisis de varianza

Análisis de covarianza multivariado

	VI_1			
	$\bar{x}_{1_{VD1}}$	$\bar{x}_{1_{VD2}}$	$\bar{x}_{2_{VD1}}$	$\bar{x}_{2_{VD2}}$
VI_2	$\bar{x}_{3_{VD1}}$	$\bar{x}_{3_{VD2}}$	$\bar{x}_{4_{VD1}}$	$\bar{x}_{4_{VD2}}$
	$\bar{x}_{5_{VD1}}$	$\bar{x}_{5_{VD2}}$	$\bar{x}_{6_{VD1}}$	$\bar{x}_{6_{VD2}}$

Una/o más Variable/s continua/s que influyen sobre las VD's

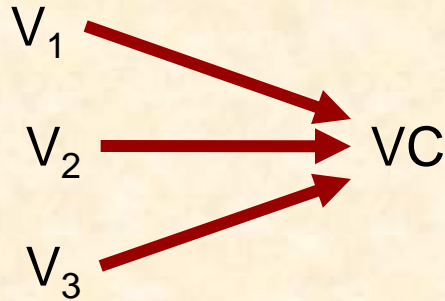
Una/o más Variable/s continua/s que influyen sobre las VD's



Estos modelos de análisis también admiten variables continuas que influyen sobre las variables dependientes (covariables).

Modelos de clasificación

Análisis de regresión lineal



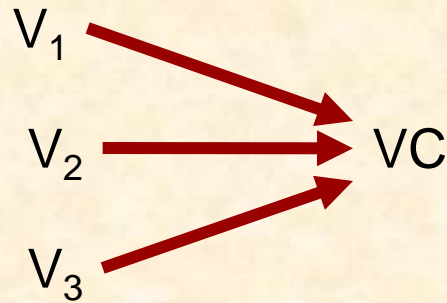
Ejemplo: se desea saber en una empresa si todas las pruebas empleadas para la selección de personal predicen el desempeño laboral de sus empleados, para ello califican el desempeño de cada uno de sus trabajadores y les aplican la batería completa de pruebas.

La finalidad de éste análisis es conocer cuáles predictores, de entre un conjunto de variables numéricas, permiten predecir la calificación que es posible obtener en una determinada escala también medida de forma numérica, sus cálculos se basan en la ecuación de la línea recta:

$$\hat{Y} = a + b_1X_1 + b_2X_2 + \dots b_nX_n$$

Modelos de clasificación

Análisis de regresión logística binaria



Ejemplo: se han medido un conjunto de variables que se supone se relacionan a la probabilidad de que ocurra una enfermedad “n”, se desea saber si éstas predicen la probabilidad de presentar la enfermedad.

La finalidad de éste análisis es conocer cuáles predictores, de entre un conjunto de variables numéricas, ordinales o dicotómicas, predicen la probabilidad de que ocurra un evento, en ese sentido, la variable a predecir es una variable dicotómica en la que se tienen un grupo que tiene la característica y un grupo de referencia que no la presenta.

Modelos de clasificación

Análisis discriminante

Variable de
Agrupación

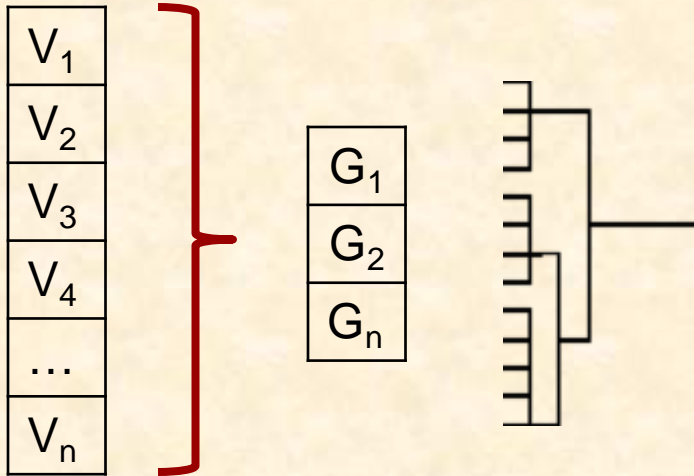
	G_1	G_2	G_3
V_1			
V_2			
V_3			
V_4			
...			
V_n			

Ejemplo: se requiere saber qué variables de la familia perfilan las diferencias existentes entre niños que viven en situación de calle (G_1), que viven con sus familias pero trabajan en la calle (G_2) o que viven con sus familias (G_3), se utilizan diferentes escalas que miden aspectos como: maltrato (V_1), comunicación (V_2), apoyo (V_3), nivel adquisitivo de la familia (V_4), entre otras (V_n).

La finalidad de éste análisis es conocer cómo un conjunto variables numéricas interactúan para predecir el pertenecer a uno de los grupos o categorías de la variable de agrupación.

Modelos de agrupación

Análisis de conglomerados o cúmulos



Ejemplo: se tienen diferentes escalas que miden aspectos como: maltrato (V_1), comunicación (V_2), apoyo (V_3), nivel adquisitivo de la familia (V_4), entre otras (V_n). Se desea conocer si éstas permiten identificar grupos diferentes.

La finalidad de éste análisis es conocer cómo un conjunto variables numéricas o binarias interactúan para separar a los sujetos en grupos.

Modelos de agrupación

Análisis Factorial

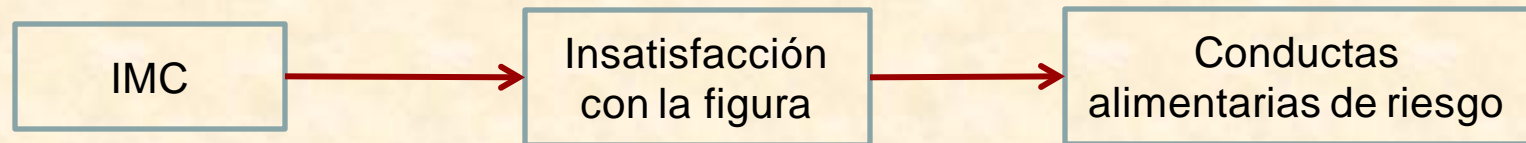
F_1	F_2	F_n
V_1		
V_2		
V_3		
	V_4	
	V_5	
	V_6	
		V_7
		V_8
		V_n

Ejemplo: se ha construido una escala para medir supervisión parental, considerando que ésta consiste en brindar límites y reglas, monitoreo o control de las actividades de los hijos y la comunicación espontánea que hacen los hijos de sus actividades. Se espera que los reactivos que conforman la escala se agrupen en estos tres factores (reglas, monitoreo y comunicación de actividades).

La finalidad de éste análisis es conocer cómo un conjunto variables se agrupan en factores. La idea central es que las variables que conceptualmente se asocian conformen entre sí un factor.

Modelos estructurales de ecuaciones

Análisis de Trayectorias (Path analysis)



Fuente: Unikel et al., 2009

$X^2 = nn.nn$, $gl = nn$, $p = n.nn$, $CFI = .nnn$, $RMSEA = 0.nn$, $IC_{90\%} RMSEA = .000 - .nnn$, $r^2 = .nn$

Figura 1. Análisis de trayectorias de la relación entre conductas alimentarias de riesgo, índice de masa corporal e insatisfacción con la figura.

- Este tipo de análisis permite evaluar la relación entre variables en diferentes niveles (antecedentes, mediadoras y resultantes) en un solo modelo.
- Las regresiones lineales están representadas por flechas unidireccionales, las correlaciones con flechas bidireccionales.
- Los diferentes niveles de relaciones se diagraman de izquierda a derecha finalizando con las variables resultantes.
- Entre los principales índices de ajuste del modelo se tienen: prueba de bondad de ajuste de X^2 (debe ser no significativa), coeficiente de ajuste CFI (debe ser mayor a .95) y el índice de error $RMSEA$ (debe ser menor a 0.05).

Modelos estructurales de ecuaciones

FASEM (Factorial structural equation model)

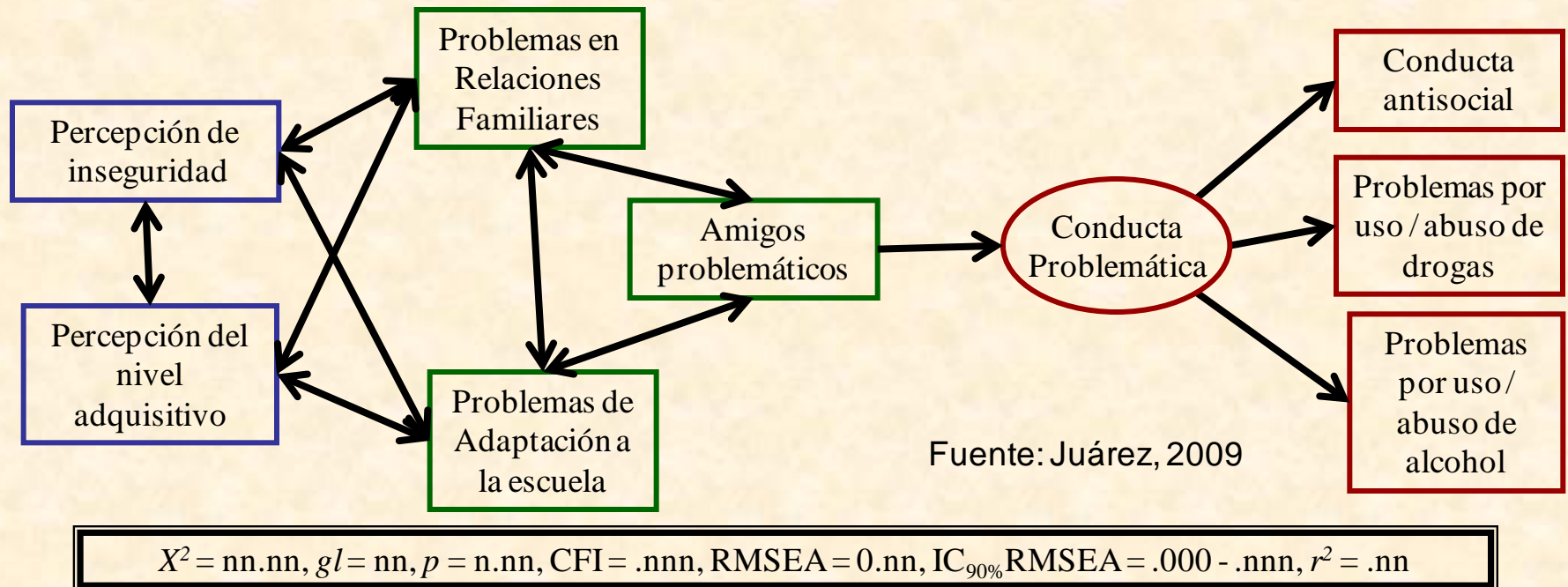


Figura 1. Relación de la conducta problemática con factores macro y micro sociales.

Similares a los modelos de trayectorias, aunque en los modelos FASEM se incluyen además factoriales confirmatorios que crean variables latentes (óvalos) y su relación con variables observadas (rectángulos).