



Universidad **Ricardo Palma**

RECTORADO
PROGRAMA DE ESPECIALIZACIÓN EN CIENCIA DE DATOS

Formamos seres humanos para una cultura de paz

TALLER DE ESTADÍSTICA PARA LA CIENCIA DE DATOS



R + Python



MÓDULO 4 : ANALISIS MULTIVARIADO I

Análisis de Correspondencias Simples



A nuestro recordado Maestro

Dr. Erwin Kraenau Espinal, Presidente de la Comisión de Creación de la Maestría en Ciencia de los Datos



TALLER DE ESPECIALIZACIÓN “STATISTICAL SCIENCE INTRODUCTION”

TALLER DE ESTADÍSTICA PARA CIENCIA DE DATOS

EXPOSITORES



José Antonio Cárdenas Garro
UNMSM

MSc in Data Science Candidate
Promotion "Erwin Kraenau Espinal"
Universidad Ricardo Palma



André Omar Chávez Panduro
UNMSM

MSc in Data Science Candidate
Promotion "Erwin Kraenau Espinal"
Universidad Ricardo Palma

**Predictive Modelling
Specialist**



Data Scientist



**Portfolio and
Consumption Analyst**



**Customer Intelligence
Analyst**



Data Analyst



Data Analyst



Correo : josecardenasgarro@gmail.com
LinkedIn : www.linkedin.com/in/jos%C3%A9-antonio-c%C3%A1rdenas-garro-599266b0

Correo : andrecp38@gmail.com /
09140205@unmsm.edu.pe
LinkedIn : www.linkedin.com/in/andr%C3%A9-ch%C3%A1vez-a90078b9



TALLER DE ESPECIALIZACIÓN "STATISTICAL SCIENCE INTRODUCTION"

« Divide las dificultades que examinas en tantas partes como sea posible , para su mejor solución»



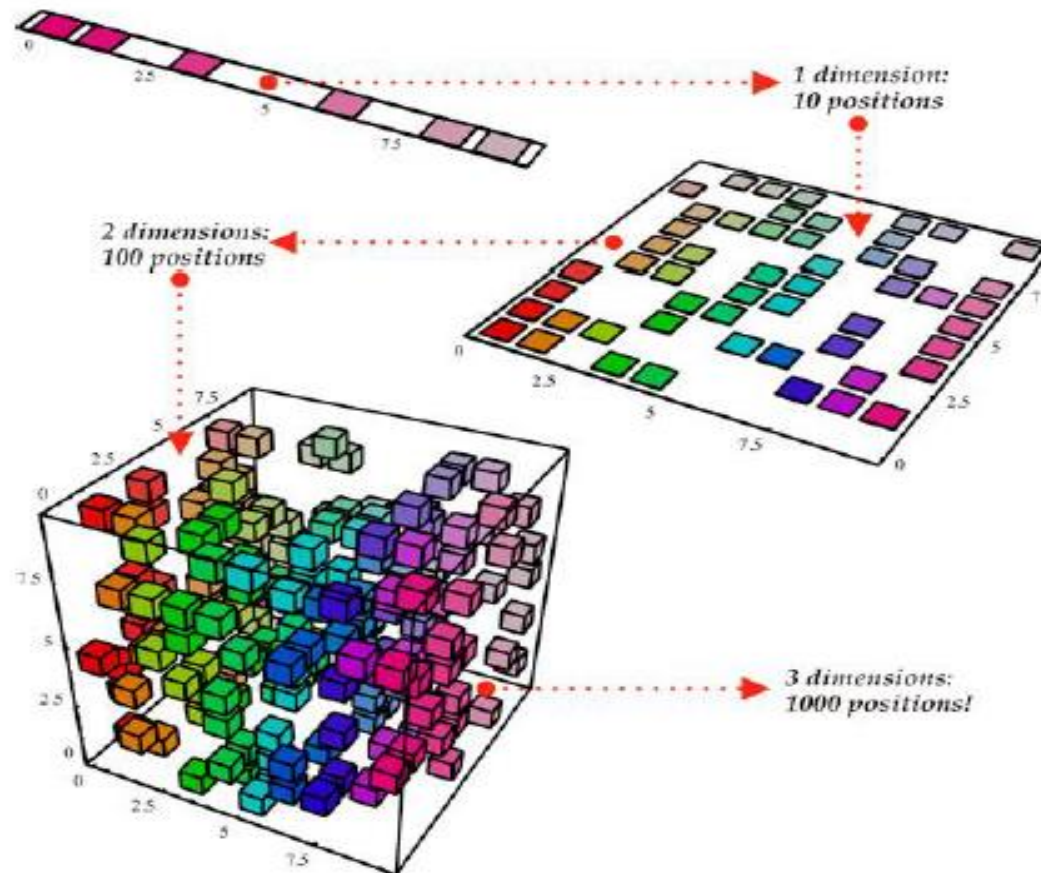
AGENDA

- Reducción de Dimensionalidad
- Análisis de Correspondencia



Reducción de Dimensionalidad

- La “Maldición de la Dimensionalidad”.



Análisis de correspondencias

Análisis de correspondencias

➤ *Objetivo*

➤ *Algoritmo*

➤ *Interpretación de resultados*

Análisis de correspondencias (objetivo)

✓ Realizar un proceso similar al análisis factorial para tablas de contingencia de variables categóricas (usualmente con muchas categorías)

Cuenta de id	vida			
ecivil	Aburrida	Excitante	Rutinaria	Total general
Casado	3,93%	48,92%	47,15%	100,00%
Divorciado	6,76%	43,92%	49,32%	100,00%
No contesta	0,00%	100,00%	0,00%	100,00%
Separado	24,00%	32,00%	44,00%	100,00%
Soltero	5,61%	54,59%	39,80%	100,00%
Viudo	15,45%	36,36%	48,18%	100,00%
Total general	6,47%	47,52%	46,01%	100,00%

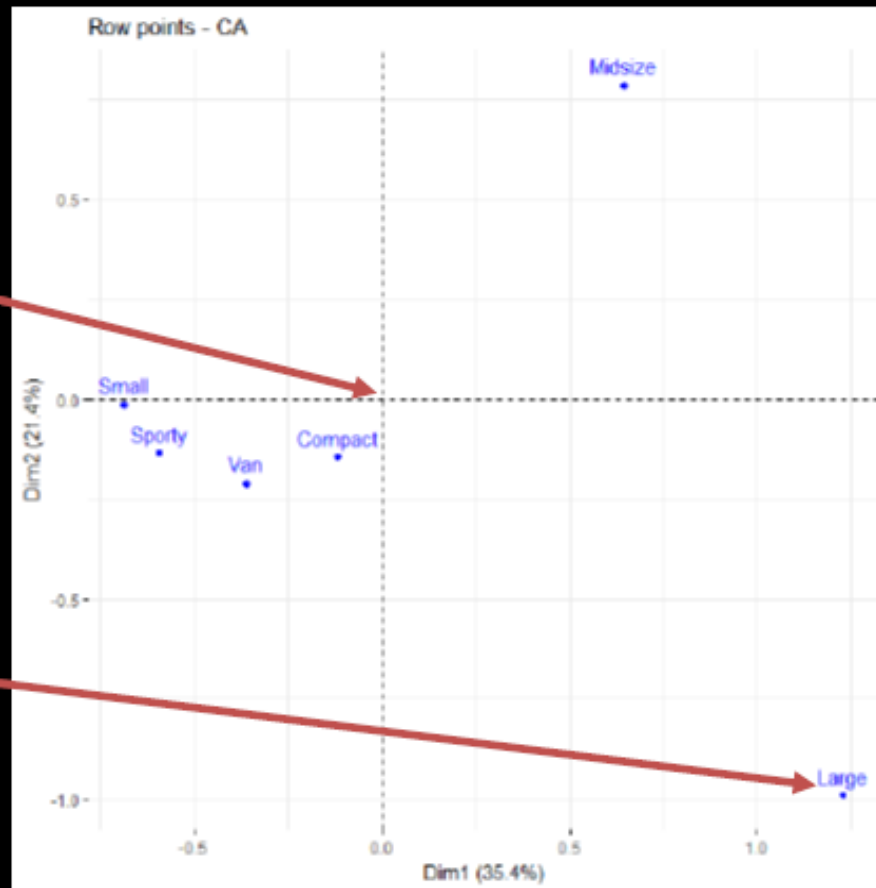
Análisis de correspondencias (algoritmo)

- ✓ Se calcula distancia chi-cuadrado entre porcentajes de fila ($\% \text{ fila } i - \% \text{ fila } j$)/tot col k
- ✓ A estas distancias se les obtienen componentes principales
- ✓ Centroide o baricentro: Fila promedio
- ✓ Inercia de la fila: total de fila por distancia al centroide al cuadrado
- ✓ La suma de todas las inercias es el chi cuadrado/n clásico

Análisis de correspondencias (interpretación I)

Centroide

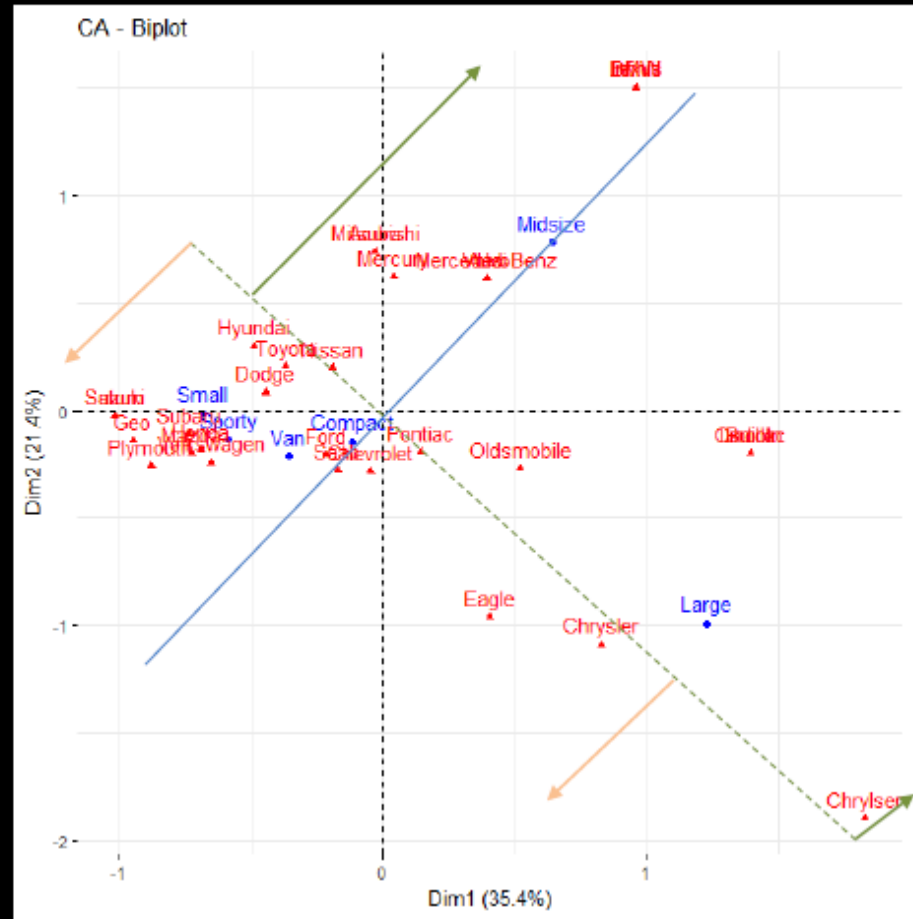
Posible alta
inercia



Análisis de correspondencias (interpretación I)

Zona verde: Más frecuente la marca de lo esperado para ese tamaño. A más distancia de la línea azul, más frecuente la marca en el tamaño de auto (absoluto).

Zona naranja: Menos frecuente la marca de lo esperado para ese tamaño. A más distancia de la línea azul, menos frecuente la marca en el tamaño de auto (absoluto).





¡Gracias!

TALLER DE ESPECIALIZACIÓN “STATISTICAL SCIENCE INTRODUCTION”