

Repaso

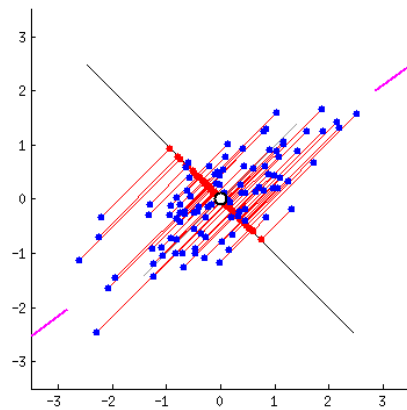
Reducción de dimensiones

Antonio Alvarez

aqalag94@gmail.com

12-Jun-2020

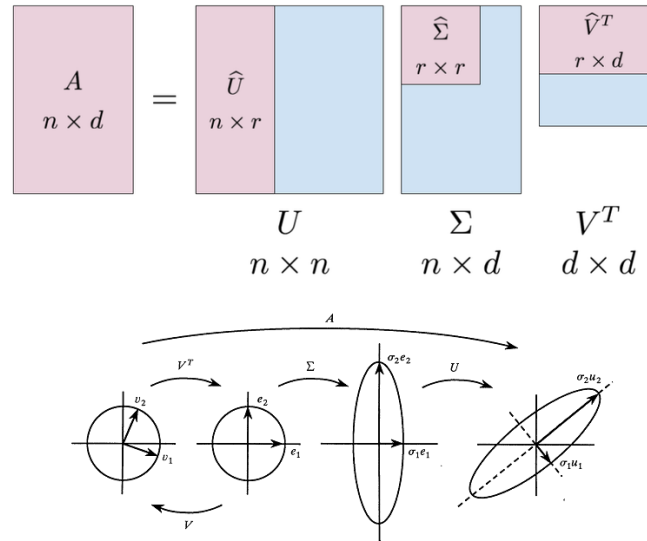
Encontrar la dirección de mayor varianza



Descomposición de Valores Singulares

- Método de Factorización para cualquier matriz rectangular

- $U \rightarrow$ componentes de las observaciones
- $\Sigma \rightarrow$ diagonal de las varianzas
- $V \rightarrow$ componentes de las variables



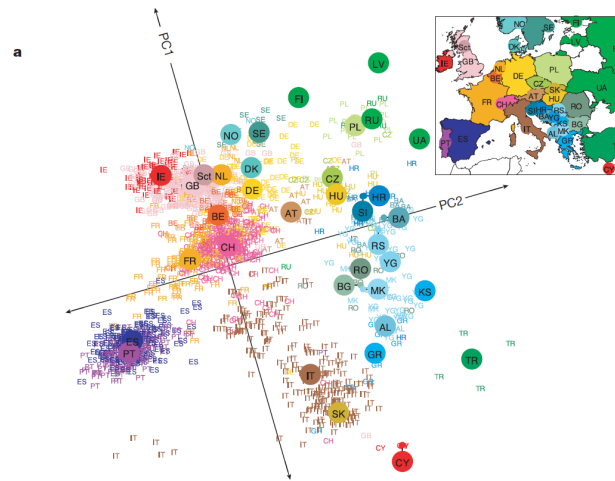
Análisis de Componentes Principales

Principal Component Analysis (PCA)

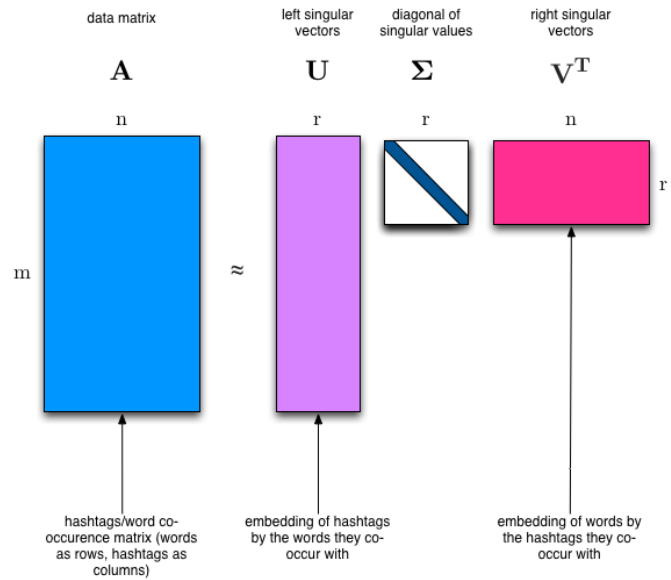
Encuentra la colección de direcciones ortogonales de máxima dispersión

Solo valores numéricos

Clásico: sensible a aberrantes



Fuente: John Novembre, et al. "Genes mirror geography within Europe." *Nature* 456 (2008)



Análisis de Correspondencia

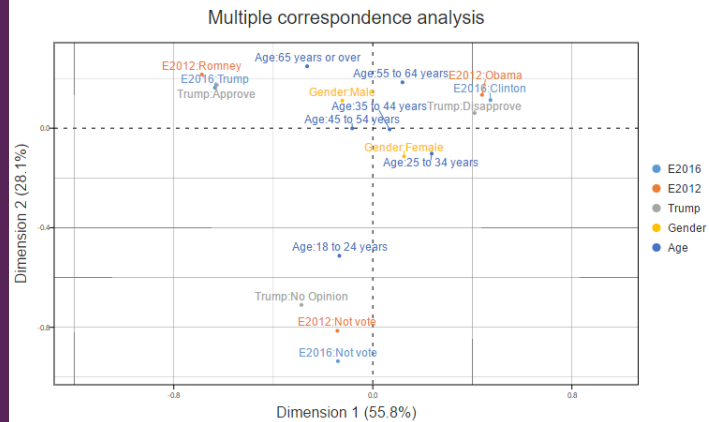
Correspondence Analysis (CA)

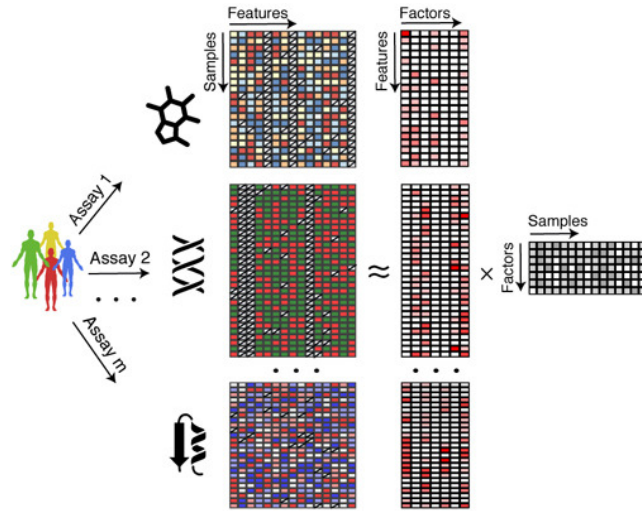
- Estudia la independencia de 2 variables categóricas
- Tablas de Contingencia
- Trata a columnas y filas de manera equivalente

Análisis de Correspondencia Múltiple

Multiple Correspondence Analysis (MCA)

- Individuos con varias variables categóricas
- Extensión de CA para varias categorías
- PCA para variables categóricas





Análisis de Factores Múltiples

Múltiple Factor Analysis (MFA)

- Individuos con varias variables numéricas/categóricas
- Considera la estructura de grupos
- Balancea la dispersión que cada grupo trae al análisis

Ayuda en Interpretaciones

Coseno al cuadrado

$$\cos^2_{i,l} = \frac{f_{i,l}^2}{\sum_l f_{i,l}^2} = \frac{f_{i,l}^2}{d_{i,g}^2}$$

Determina el nivel de representación del elemento en ese componente

Contribución

$$\text{ctr}_{i,l} = \frac{f_{i,l}^2}{\sum_i f_{i,l}^2} = \frac{f_{i,l}^2}{\lambda_l}$$

Dependencia del componente al elemento para su existencia

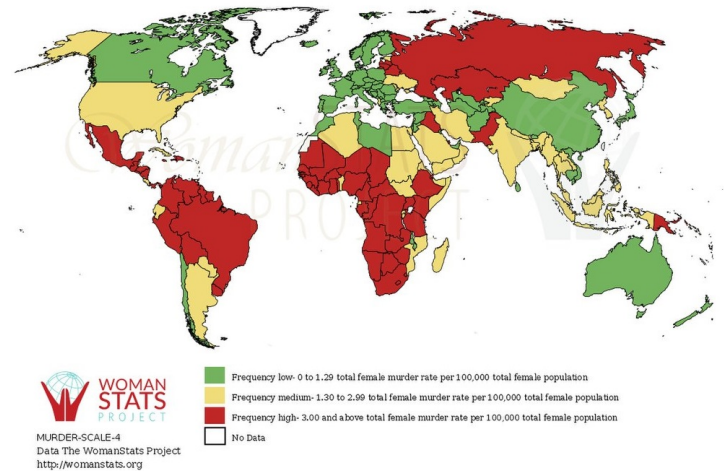
Otros métodos

- *Non-Negative Matrix Factorization*
- *t-distributed Stochastic Neighbour Embedding*
- *Robust PCA*

Violencia de Género

Preámbulo

- Definimos violencia de género como actos primariamente o exclusivamente cometido en contra de mujeres o niñas



- Estos muchas veces son considerados como crímenes de odios, causados por sentimientos de derecho, superioridad o misoginia

ENVIGMU 2019

Encuesta Nacional sobre Relaciones Familiares y Violencia de Género Contra las Mujeres

- Recoge testimonios de mujeres solteras, casadas y separadas a través de cuestionarios
- Las preguntas se pueden agrupar en temas: 5 ámbitos y 5 tipos de violencia

Detalles

- Encuesta por muestreo probabilístico
- Mujeres de 15 años o mas
- 20448 viviendas, 12057 mujeres
- Estimación nacional, urbana/rural, provincial
- Violencia a lo largo de su vida y en el ultimo año

4 ámbitos, 6 grupos

```
library(readr)
vio_df <- read_delim("data/violencia.csv", ";")[-1]

# Pesos de la encuesta
vio_p <- vio_df %>% select(pesos)

dim(vio_df)
```

```
## [1] 12057    72
```

16 Preguntas Educativas, 18 Laborales, 14 Sociales, 22 Familiares, 1 edad, 1 estado civil

Análisis de la encuesta



Explorar la estructura de los datos

- Reportar la varianza de los 10 primeros componentes
- Describir la nube de individuos y de preguntas
- Interpretar el primer y el segundo componente
- Limitaciones del análisis
- Siguiendo pasos



Materiales

- Cuestionario (`violencia.csv`)
- Etiquetas de las preguntas (`etiquetas.xlsx`)
- Link
- Pagina del INEC



Entregables

- Script de R
- Por cada enunciado de exploración, tener código que responde el enunciado + el comentario que refleja la respuesta