

PCA

Diego Isau Barranco Herrera

Principal Component Analysis (PCA)

El PCA es una técnica de reducción de dimensión. A través de combinaciones lineales de las variables involucradas podemos reducir la cantidad de estas, siempre y cuando no se pierda información, esto es, tener como mínimo una varianza acumulada del 70% entre todos los componentes

Población USA

Tenemos 19 variables independientes, vamos a separar por año para realizar el PCA

```
poblacion <- read_xlsx("/var/home/diegob/Covid.xlsm")
```

PCA para el año 2000

```
poblacion_2000 <- poblacion %>% select(c(2,3,5,7,9,11,13,15,17,19))  
  
# normalizando datos  
poblacion_2000_normal <- scale(poblacion_2000)  
  
# Correlación  
det(cor(poblacion_2000_normal)) # La correlación tiende a cero
```

```
[1] 6.876395e-41
```

```
#psych::cor.plot(poblacion_2000_normal)
```

Ahora vamos a verificar el Factor de adecuación muestral de Kaiser, el cual nos indica si las variables son aptas para realizar el PCA

```
psych::KMO(poblacion_2000_normal)
```

```
Error in solve.default(r) :
```

```
System ist für den Rechner singulär: reziproke Konditionszahl = 1.23692e-18
```

```
Kaiser-Meyer-Olkin factor adequacy
```

```
Call: psych::KMO(r = poblacion_2000_normal)
```

```
Overall MSA = 0.5
```

```
MSA for each item =
```

| | |
|--|-----|
| Census Resident Total Population - AB:Qr-1-2000 | 0.5 |
| Resident Total Population Estimate - Jul-1-2000 | 0.5 |
| Net Domestic Migration - Jul-1-2000 | 0.5 |
| Federal/Civilian Movement from Abroad - Jul-1-2000 | 0.5 |
| Net International Migration - Jul-1-2000 | 0.5 |
| Period Births - Jul-1-2000 | 0.5 |
| Period Deaths - Jul-1-2000 | 0.5 |
| Resident Under 65 Population Estimate - Jul-1-2000 | 0.5 |
| Resident 65 Plus Population Estimate - Jul-1-2000 | 0.5 |
| Residual - Jul-1-2000 | 0.5 |