

S1_LC_Taller_PCA

February 24, 2022

1 Caso-taller: PCA para estudiar la dimensionalidad de la calidad de vida en las ciudades

En el Almanaque Calificado de Lugares, Boyer y Savageau calificaron 329 ciudades de acuerdo con los siguientes nueve criterios:

- Clima y Terreno
- Alojamiento
- Cuidado de la salud y el medio ambiente
- Crimen
- Transporte
- Educación
- Las artes
- Recreación
- Economía

Con tantas variables disponibles, la matriz de dispersión puede resultar muy grande para estudiarla e interpretarla adecuadamente. Habría demasiadas correlaciones por pares entre las variables para considerar. Para interpretar los datos de una forma más significativa es necesario reducir el número de variables a unas pocas dimensiones, y para ello en este ejercicio recurriremos al análisis de componentes principales.

Este taller tiene como objetivos generales que usted sea capaz de:

1. Reconocer las características y el funcionamiento del análisis de componentes principales.
2. Construir e implementar el análisis de componentes principales.
3. Interpretar los resultados del análisis de componentes principales para determinar el número adecuado de componentes a utilizar en la resolución de un problema en particular.

y en particular que usted sea capaz de:

1. Realizar un análisis de componentes principales utilizando Álgebra Lineal en **Python**.
2. Evaluar cuántos componentes principales se necesitan.
3. Interpretar los pesos de los componentes principales y describir una dimensión con un puntaje alto o bajo.
4. Determinar cuándo un análisis de componentes principales debe basarse en la matriz de varianza-covarianza o en la matriz de correlación.
5. Utilizar los pesos de los componentes principales en análisis posteriores.

1.1 Instrucciones generales

1. Para desarrollar el *cuaderno* primero debe descargarlo junto a los archivos de soporte.
2. Para responder cada inciso deberá utilizar el espacio debidamente especificado.
3. La actividad será calificada sólo si sube el *cuaderno* de jupyter notebook con extensión `.ipynb` en la actividad designada como “entrega calificada por el personal”.
4. El archivo entregado debe poder ser ejecutado localmente por el tutor. Sea cuidadoso con la especificación de la ubicación de los archivos de soporte, guarde la carpeta de datos en el mismo `path` de su *cuaderno*, por ejemplo: `datos/archivo_de_soporte.csv`.

1.2 Criterios de Evaluación

1.3 Desarrollo

Carga de datos y Análisis preliminar de las variables

1. Carga de archivo de soporte. En la carpeta de `datos` se encuentra el archivo `lugares.csv` cargue estos datos en su *cuaderno*.

[1]: `# Utilice este espacio para escribir el código.`

(Utilice este espacio para describir su procedimiento.

2. Análisis descriptivo de las variables. Para el análisis descriptivo haga uso de correlogramas, histogramas y estadísticas descriptivas. Indique además cuantas gráficas de dispersión bivariadas debería examinar con este número de variables. Tenga en cuenta que para las variables, a excepción de la alojamiento y crimen, cuanto mayor sea el puntaje, mejor. Por el contrario; para alojamiento y crimen, mientras más bajo sea el puntaje, mejor. También, mientras algunas comunidades podrían calificar mejor en las artes, otras podrían calificar mejor en otras áreas como tener una tasa de criminalidad más baja y buenas oportunidades educativas.

[2]: `# Utilice este espacio para escribir el código.`

(Utilice este espacio para describir el procedimiento, análisis y conclusiones).

3. Cálculo y selección de Componentes Principales. Utilizando Álgebra Lineal, calcule los eigen valores y los eigen vectores. Explique la conexión entre los eigen valores, los eigen vectores, y los componentes principales. Examine los valores propios para determinar cuántos componentes principales se deben considerar. Explique el criterio que utilizó para la elección y justifique su respuesta.

[3]: `# Utilice este espacio para escribir el código.`

(Utilice este espacio para describir el procedimiento, análisis, y conclusiones).

4. Interpretación de los Componentes Principales Logró reducir la dimensión en el ejercicio anterior? Si lo hizo, los pesos obtenidos pueden identificar dimensiones menores de los datos? Dé una interpretación a estas dimensiones, para ello puede utilizar gráficas (ej. `biplot`), tablas, etc.

[4]: *# Utilice este espacio para escribir el código.*

(Utilice este espacio para describir el procedimiento, análisis, y conclusiones).

5. Detección de ciudades inusuales A partir de los componentes seleccionados, identifique si existen ciudades inusuales o atípicas. Explique claramente qué procedimiento utilizó para identificarlas e interprete los resultados incluyendo qué entiende por ciudad inusual o atípica.

[5]: *# Utilice este espacio para escribir el código.*

(Utilice este espacio para describir el procedimiento, análisis, y conclusiones).

6. Estandarice los datos y repita los pasos anteriores. Repita los pasos anteriores con los datos estandarizados. Compare los resultados que obtuvo previo a, y luego de estandarizar los datos. Los resultados cambian significativamente? Justifique su respuesta.

[6]: *# Utilice este espacio para escribir el código.*

(Utilice este espacio para describir el procedimiento, análisis, y conclusiones).