



Data&Analytics
INNOVACIÓN Y TECNOLOGÍA



Data&Analytics
INNOVACIÓN Y TECNOLOGÍA

M.Sc. Angelo Jonathan Diaz Soto



Análisis de Componentes Principales

2025

Data Science – Business Intelligence – Big Data – Machine Learning – Artificial Intelligence – Innovation and

Contenido



- ✓ Conceptos básicos
- ✓ Cálculo de las PCA
- ✓ Criterios para

determinar el número de PCA ✓ Interpretación
de los PCA

✓ Referencias bibliográficas

(+51) 976 760 www.datayanalytics.com info@datayanalytics.com

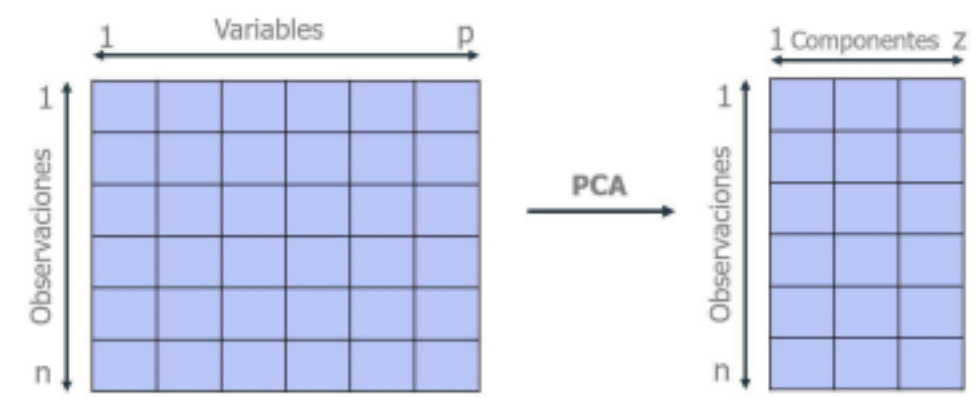
Introducción



✓ El análisis de
(Principal Component
reducción de

componentes principales
Analysis PCA) **es un método de
dimensionalidad** que permite

simplificar la complejidad de espacios con múltiples dimensiones a la vez que conserva su información.



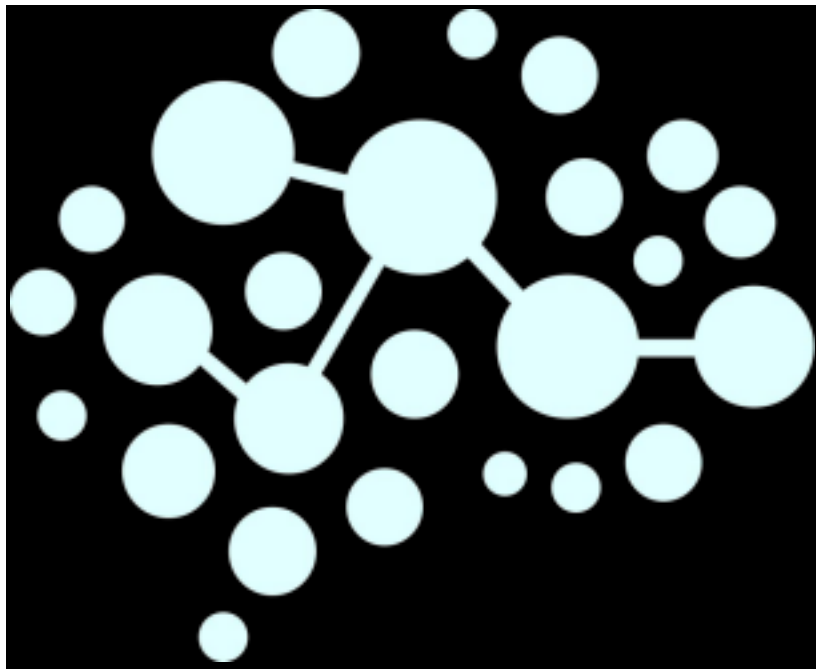
(+51) 976 760 www.datayanalytics.com info@datayanalytics.com

Introducción

$$\begin{pmatrix} \mathbf{x}^1 & \mathbf{x}^p \end{pmatrix}$$
$$\mathbf{x}^j$$
$$\mathbf{1}_1$$

$$\begin{pmatrix} \mathbf{x}^1 & \mathbf{x}^p \end{pmatrix} \mathbf{X} = \begin{pmatrix} \mathbf{x}^1 & \mathbf{x}^p \end{pmatrix}$$
$$\mathbf{x}^j$$
$$\mathbf{x}^p$$

$x_1 \dots x_j$



\mathcal{I}_n

$\mathbf{Y} = \begin{bmatrix} y^1 \\ \vdots \\ y^n \end{bmatrix}$

\mathbf{y}_j
 \mathbf{y}^j
 \mathbf{y}^q

Objetivo:

Definir q nuevas variables,
 $q < p$, reteniendo el máximo de información de \mathbf{X} .

[illegible]

(+51) 976 760 www.datayanalytics.com info@datayanalytics.com



Introducción

✓ El análisis de PCA es una

componentes principales o de las técnicas de

aprendizaje no supervisado, las cuales suelen aplicarse como parte del análisis exploratorio de los datos.

✓ Una de las aplicaciones de PCA es la reducción de dimensionalidad ✓

El PCA también sirve como herramienta para la visualización de datos

✓ El PCA **no requiere la suposición de normalidad multivariante** de los datos

(+51) 976 760 www.datayanalytics.com info@datayanalytics.com

Explicación geométrica del PCA







(+51) 976 760 www.datayanalytics.com info@datayanalytics.com

Cálculo de las PCA



✓ Como se
una
de las

ha dicho anteriormente, los
componentes principales son
combinación lineal normalizada
variables originales de un set

de datos.

- ✓ Los componentes principales son **autovectores** que se toman de la matriz de correlaciones
- ✓ Generalmente, se podrán obtener tantas componentes principales distintas como variables disponibles.
- ✓ La elección se realiza de manera que la primera componente principal sea la que **mayor varianza** recoja; la segunda debe recoger la máxima variabilidad no recogida por la primera, y así sucesivamente, eligiendo un número que recoja un porcentaje suficiente de varianza total.

(+51) 976 760 www.datayanalytics.com info@datayanalytics.com

Cálculo de las PCA



Algebraicamente, las

componentes principales son una combinación lineal de las p variables aleatorias originales X_1, X_2, \dots, X_p y geoméricamente esta combinación lineal representa la elección de un nuevo sistema de coordenadas obtenidas al rotar el sistema original.

- ✓ Estos nuevos ejes representan la dirección de máxima variabilidad.
- ✓ Por lo tanto el ACP permite describir la estructura e interrelación de variables originales consideradas simultáneamente, determinando q combinaciones lineales de las p -variables originales que expliquen la mayor parte de la variación total, y de esta forma resumir y reducir los datos.

(+51) 976 760 www.datayanalytics.com info@datayanalytics.com

Cálculo de las PCA







Cálculo de las PCA



$$Y_p = a_p X = a_{p1} X_1 + a_{p2} X_2 + \dots + a_{pp} X_p$$

Entonces las variables Y_1, Y_2, \dots, Y_p son las componentes principales, las mismas que no están correlacionadas entre sí, son ortonormales entre ellas y además se cumple que:

$$Var(Y) = \lambda$$

(+51) 976 760 www.datayanalytics.com info@datayanalytics.com

Cálculo de las PCA



porcentaj

por:

componente principal o su explicación está dado



$$\sum_{i=1}^p \lambda_i$$

- ✓ El porcentaje total de la varianza contenida por las **q** primeras componentes principales se define así:

$$\frac{\sum_{i=1}^q \lambda_i}{\sum_{i=1}^p \lambda_i}$$

(+51) 976 760 www.datayanalytics.com info@datayanalytics.com



Criterios para determinar el número de PCA

- ✓ En general, el obtener el principales a autovalores en **75%** de la muestra.



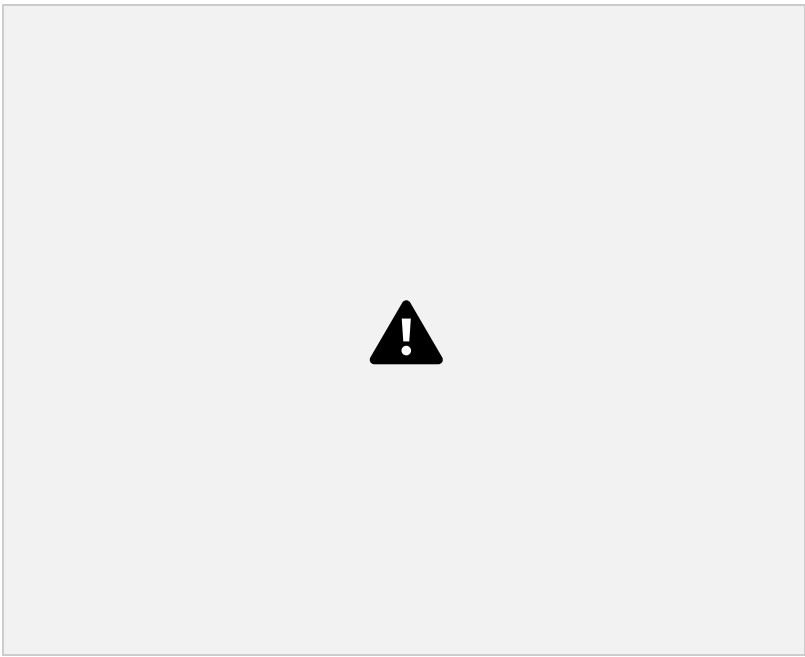
criterio más sencillo para número **m** de componentes retener debe ser tal que los conjunto explican **más del** información total de la

- ✓ **Gráfico de sedimentación.** En este gráfico en el eje Y se representan los valores propios o raíces características y en el eje X el número de componentes principales correspondientes a cada valor propio en orden decreciente, de acuerdo a este gráfico **se retienen aquellas componentes que se encuentran antes de que el gráfico presente un "quiebre" o "codo"**.

(+51) 976 760 www.datayanalytics.com info@datayanalytics.com

Criterios para determinar el número de PCA

- ✓ Gráfico de sedimentación.



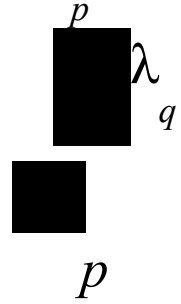


(+51) 976 760 www.datayanalytics.com info@datayanalytics.com

Criterios para determinar el número de PCA



ca. según este criterio se retienen aquellas componentes



✓ y se seleccionan aquellas componentes cuya raíz característica excede de la media de las raíces características.

Valor propio (autovector)



Los valores propios (también llamados valores característicos o raíces latentes) son las varianzas de los componentes principales.



Usted puede utilizar el tamaño del valor propio para determinar el número de componentes principales. Conserve los componentes principales con los valores propios más grandes. Por ejemplo, según el criterio de Kaiser, **se usan solo los componentes principales con valores propios que son mayores que 1.**



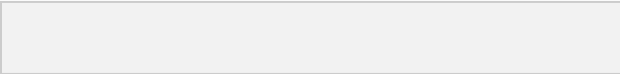


(+51) 976 760 www.datayanalytics.com info@datayanalytics.com

Valor propio (autovector)







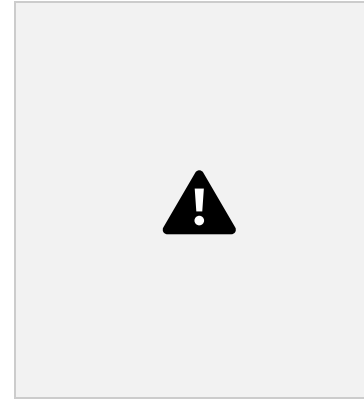
(+51) 976 760 www.datayanalytics.com info@datayanalytics.com



Componentes principales (PC)

Para interpretar cada componente principal, examine la magnitud y la

dirección de los coeficientes de las variables originales. **Cuanto mayor sea el valor absoluto del coeficiente, más importante será la variable correspondiente** en el cálculo del componente. Qué tan grande debe ser el valor absoluto de un coeficiente para ser considerado importante es subjetivo. Utilice su conocimiento especializado para determinar en qué nivel el valor de correlación es importante.



(+51) 976 760 www.datayanalytics.com info@datayanalytics.com

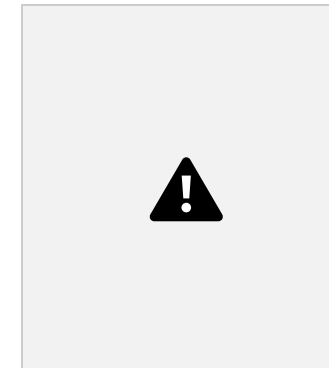


Componentes principales (PC)

En estos resultados, el **primer**

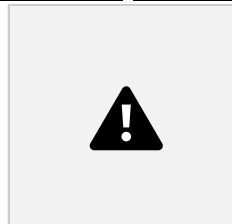
componente

principal tiene asociaciones positivas grandes con Edad, Residencia, Empleo y Ahorros. Usted puede interpretar este componente principalmente como una medición de la estabilidad financiera a largo plazo de un solicitante. El **segundo componente** tiene asociaciones negativas grandes con Deudas y Tarjeta Crédito, así que este componente mide principalmente el historial crediticio de un solicitante. El **tercer componente** tiene asociaciones negativas grandes con ingresos, educación y tarjeta crédito, así que este componente mide principalmente las calificaciones académicas y de ingresos de un solicitante.



(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com

Puntuaciones



Las **puntuaciones son combinaciones lineales de los datos que se determinan por los coeficientes de cada componente principal.**

Para obtener la puntuación de una observación, sustituya sus valores en la ecuación lineal del componente principal. Si utiliza la matriz de correlación, usted debe estandarizar las variables para obtener la puntuación correcta de los componentes cuando usa la ecuación lineal.



(+51) 976 760 www.datayanalytics.com info@datayanalytics.com

Puntuaciones

En estos resultados, la puntuación del primer componente principal se puede calcular a partir de los datos estandarizados utilizando los coeficientes indicados en PC1:



PC1 = 0.314 Ingresos + 0.237
Educación +
0.484 Edad + 0.466 Residencia +
0.459 Empleo + 0.404 Ahorros - 0.067
Deudas -
0.123 Tarj_crédito



(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com



Referencias bibliográficas

✓ Marzal Varó,
con

✓ Guagliano,
Python I:



A., & Gracia Luengo, I. (2009).
Introducción a la programación
Python. Universitat Jaume I.

C. (2019). Programación en
Entorno de Programación-

- ✓ Menezes, N. N. C. (2017). Introducción a la programación con Python: Algoritmos y lógica de programación para principiantes. Novatec Editora.
- ✓ https://rpubs.com/Cristina_Gil/PCA
- ✓ <https://support.minitab.com/es-mx/minitab/18/help-and-how-to/modeling-statistics/multivariate/how-to/principal-components/interpret-the-results/all-statistics-and-graphs/>





(+51) 976 760 www.datayanalytics.com info@datayanalytics.com