**MÉTODOS ESTADÍSTICOS**

|  |  |
| --- | --- |
| Nombre(s): Equipo 4:   * Diana Zepeda Martínez * José Juan García Romero | |
| Nº 24 | Realizar los ejercicios en Rstudio aplicando la teoría de PCA  Incluir la descripción de nuevas funciones y una interpretación general. |
| a) Preparación de la base de datos  datos <- Base\_de\_datos\_de\_colombia\_tv\_internet[2:9]  row.names(datos)<-Base\_de\_datos\_de\_colombia\_tv\_internet$DEPTO | |
|  | |

|  |
| --- |
| b) Análisis de correlación |
|  |
| c) ACP con función prcomp y con estandarización de datos |
|  |
| c.1) mostrar el porcentaje de varianza acumulada de los CP |
|  |

|  |
| --- |
| c.2) Gráfico de sedimentación de varianzas |
|  |
| c.3) Gráfico de variables e individuos |
|  |
| d) ACP con función PCA creando gráficos de variables e individuos |
|  |
| d.1) Gráfica de los porcentajes de varianzas de los componentes |
|  |
| d.2) Gráfica de las variables de los componentes |
|  |
| d.3) Gráfica de los individuos de los componentes |
|  |
| d.4) Gráfica conjunta (variables e individuos) los componentes |
|  |
| d.5) Gráficas de la calidad que aportan las variables a los CP |
| d.5-1) Tabla de correlación-calidad variables |
|  |
| d.5-2) Histograma |
|  |
| d.5-3) Variables-PCA |
|  |
| d.6) Gráficas de la calidad que aportan los individuos a los PCA |
| d.6-1) Tabla de correlación-calidad individuos |
|  |
| d.6-2) Histograma |
|  |

|  |
| --- |
| d.6-3) Individuos-PCA |
|  |
| d.7) Gráficas de la contribución que aportan las variables a los PCA |
| d.7-1) Tabla de correlación-contribución variables |
|  |
| d.7-2) Histograma para variables - PC1 y PC2 |
|  |
| d.7-3) Variables-PCA |
|  |
| d.8) Gráficas de la contribución que aportan las variables a los PCA |
| d.8-1) Tabla de correlación-contribución variables |
|  |
| ​ d.8-2) Histograma para individuos - PC1 y PC2 |
|  |
| ​ d.8-3) Individuos-PCA |
|  |
| e) Obtener clusters jeráquicos con PCA |
|  |

|  |
| --- |
| e.1) Clúster de individuos |
|  |
| e.2) Determinar las variables que describe más los clusters |
|  |
| e.3) Determinar las variables más significativas de cada cluster |
|  |
| e.4) Determinar los componentes que más se reflejan en cada clúster |
|  |
| e.5) Individuos que más participan en cada clúster |
|  |
| f) Descripción de nuevas funciones |
| **Read\_excel:** Lee archivos Excel para leer las bases de datos que contienen.  **cor:** Muestra la correlación de datos.  **summary:** Muestra un resumen estadístico.  **biplot:** Son un tipo de gráfico exploratorio usado en Estadística.  **PCA:** Se utiliza para reducir el número de variables de forma que pasemos a tener el mínimo número de nuevas variables y que representen a todas las antiguas variables de la forma más representativa.  **fviz\_eig:** Representa por columnas los eigenvalues para cada una de las dimensiones  **fviz\_pca\_var:** Representa las variables sobre las dos primeras componentes principales.  **fviz\_pca\_biplot:** Genera un biplot.  **Corrplot:** Podemos visualizar las correlaciones entre las varibles mediante un correlograma que genera.  **fviz\_cos2:** Visualiza la calidad de representación de filas y columnas.  **fviz\_contrib:** Representa la contribución de filas y columnas de un pca.  **fviz\_pca\_ind:** Representación de observaciones sobre componentes principales.  **fviz\_cluster:** Visualiza los resultados de la agrupación en clústeres. |
| g) Interpretación general de los resultados |
| El Principal Component Analysis (PCA) es un método estadístico que permite simplificar la complejidad de espacios muestrales con muchas dimensiones a la vez que conserva su información, este método permite simplificar el análisis de grandes cantidades de datos asi mismo, podemos eliminar algunas de las variables con menos utilidad manteniendo la parte más importante todas las variables utilizadas.  Una de las aplicaciones de PCA es la reducción de dimensionalidad (variables), perdiendo la menor cantidad de información (varianza). |