

PROYECTO

1. Conjunto de Datos

La matriz de datos se deja como una elección personal, inclusive se pueden utilizar diferentes bases para desarrollar cada una de las etapas del proyecto. No olvide que las variables seleccionadas deben responder a un objetivo en particular y a las características de la técnica multivariante que se vaya aplicar.

En lo que siguen se dan las directrices para aplicar las técnicas multivariantes vistas en clases utilizando el Software Estadístico de su preferencia, una vez desarrollado el trabajo resumir los resultados obtenidos en un informe. Si algún tipo de código fuese necesario, éste debe ser incluido en un apéndice, si el código fuese una o dos líneas incluirlo en el informe, ojo que no se aceptarán informes que sean solo corridas o salidas que proporcionan los softwares estadísticos. Finalmente, es bueno aprender a seleccionar aquello que realmente es importante, redactarlo, darle un formato personal, por otro lado no olvidar que la interpretación personal de los resultados, todos estos elementos son sumamente importantes en todo el desarrollo del trabajo.

2. PRIMERA FASE:

Análisis de Componentes Principales

2.1. Poblaciones Bivariantes

En este apartado utilice una base de datos conformada por dos variables altamente correlacionadas

1. Construir los Estadígrafos Multivariantes e interpretar los mismos.
2. Realizar un análisis gráfico de los datos
3. Determinar si los datos provienen de una distribución normal bi-variante.

4. Construir la región elipsoidal con un 95 % de confianza, además determine el porcentaje de vectores de observación que se hallan dentro de la región elipsoidal.
5. Aplicar las pruebas de independencia y equicorrelación
6. Construir las componentes Principales
7. Estimar las componentes Principales y graficar las mismas.
8. Utilizando las estimaciones verificar si los supuestos del modelo de Componentes Principales se cumplen

2.2. Poblaciones Multivariantes

1. Un análisis descriptivo de los datos, se pueden utilizar elementos de la estadística univariante además de algunas representaciones gráficas propias del análisis multivariante.
2. Determinar si los datos provienen de una población normal multivariante
3. Aplicar la prueba de independencia
4. Construir las componentes asociadas al conjunto de datos
5. Calcular e interpretar ρ_{Y_i, X_j} , para $i, j = 1, 2, \dots, p$.
6. Seleccionar las componentes necesarias para representar los datos, para tal efecto aplicar los criterios:
 - Porcentaje de la Varianza
 - Promedio de los eigen valores
 - Scree Plot
 - Prueba de raíces retenidas (Suponga que los datos provienen de una población normal multivariante).

Emitir una conclusión sobre el número de componentes m que deben ser retenidas.

7. Para detectar observaciones sospechosas y evidenciar la normalidad de los datos, graficar las componentes estimadas de la muestra, es decir que se pueden construir los Q-Q plots (permiten identificar puntos atípicos y un comportamiento lineal indica normalidad en la componente), además de los diagramas de dispersión entre pares de componentes que permiten identificar estructuras elipsoidales. Aplicar estas ideas en el conjunto de datos.

8. Monitorear la calidad de los datos utilizando la relación:

$$\frac{\hat{y}_1^2}{\hat{\lambda}_1} + \frac{\hat{y}_2^2}{\hat{\lambda}_2} + \dots + \frac{\hat{y}_m^2}{\hat{\lambda}_m} \leq \chi_{m, \alpha}^2, \quad m : \text{nro. raíces retenidas}$$

9. Utilizar el gráfico T^2 para determinar si no existe información significativa en las componentes descartadas.

$$T^2 = \frac{\hat{y}_{m+1}^2}{\hat{\lambda}_{m+1}} + \frac{\hat{y}_{m+2}^2}{\hat{\lambda}_{m+2}} + \dots + \frac{\hat{y}_p^2}{\hat{\lambda}_p}$$