

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: Comparación de Medias Multivariantes

2.1. Poblaciones Bivariantes

1. Realizar el análisis descriptivo de los datos (analítico y gráfico).
2. Aplique el análisis de conglomerados, emita conclusiones al respecto.
3. Comparar los vectores de medias considerando matrices de covarianzas iguales.
4. Compare las medias considerando varianzas iguales, aplique el enfoque univariante.
5. Comparar los vectores de medias considerando matrices de covarianzas diferentes.
6. Compare las medias considerando varianzas diferentes, aplique el enfoque univariante.

7. Emita conclusiones con relación a los resultados obtenidos desde los puntos de vista uni y multivariante.

2.2. Poblaciones Multivariantes - MANOVA de una vía

1. Realizar el análisis descriptivo de los datos (analítico y gráfico).
2. Aplique el análisis de conglomerados, emita conclusiones al respecto.
3. Construir las matriz B y W utilizando las variables seleccionadas.
4. Comparar los vectores de medias considerando matrices de covarianzas iguales.
5. Compare las medias considerando varianzas iguales, aplique el enfoque univariante.
6. Construir las tablas del ANOVA para comparar las medias, desde el punto de vista univariante.
7. Emita conclusiones con relación a los resultados obtenidos desde los puntos de vista uni y multivariante.

2.3. Poblaciones Multivariantes - MANOVA de dos vías

1. Realizar el análisis descriptivo de los datos (analítico y gráfico).
2. Aplique el análisis de conglomerados, emita conclusiones al respecto.
3. Construir las matrices suma de cuadrados y productos cruzados
4. Construir la tabla del MANOVA para comparar los factores y sus interacciones.
5. Construir las tablas del ANOVA para comparar los factores y sus interacciones, desde el punto de vista univariante.
6. Emita conclusiones con relación a los resultados obtenidos desde los puntos de vista uni y multivariante.

2.4. Análisis de Perfiles

1. Realizar el análisis descriptivo de los datos (analítico y gráfico).
2. Aplicar la Prueba de Perfiles Paralelos para dos Poblaciones Normales
3. Aplicar la Prueba de Perfiles Coincidentes dado que los perfiles son paralelos.
4. Emita conclusiones con relación a los resultados obtenidos.

3. SEGUNDA FASE: Discriminación y Clasificación

3.1. Separación y Clasificación para dos poblaciones

1. Aplicar la prueba inferencial para comparar dos vectores de medias ($H_0 : \mu_1 = \mu_2$)
2. Aplicar la prueba inferencial para comparar matrices de covarianzas ($H_0 : \Sigma_1 = \Sigma_2$)
3. Construir las reglas de clasificación lineal o cuadrática para poblaciones normales y aplicarlas en el conjunto de datos
4. Calcular las probabilidades de clasificaciones erróneas
5. Calcular la tasa de error aparente utilizando la matriz de confusión y el procedimiento de Holdout.
6. Construir las funciones discriminantes de Fisher
7. Clasificar los datos utilizando la regla de Fisher y luego calcular la tasa de error aparente

3.2. Clasificación con varias poblaciones

1. Aplicar la prueba inferencial para comparar k vectores de medias ($H_0 : \mu_1 = \dots = \mu_k$)
2. Aplicar la prueba para comparar matrices de covarianzas ($H_0 : \Sigma_1 = \dots = \Sigma_k$)

3. Construir las reglas de clasificación lineal o cuadrática para poblaciones normales y aplicarlas en el conjunto de datos
4. Calcular la tasa de error aparente utilizando la matriz de confusión
5. Construir las funciones lineales de Fisher
6. Clasificar los datos utilizando la regla de Fisher y luego calcular la tasa de error aparente

4. TERCERA FASE: Análisis de Correlación Canónica

1. Determinar del conjunto de variables de su base de datos, un subconjunto de variables $x = (x_1, x_2, \dots, x_q)$ y otro subconjunto de variables $y = (y_1, y_2, \dots, y_p)$. La selección de ningún modo es aleatoria, por lo cual debe apelar a su criterio personal para identificar estos subconjuntos, considerando que puede existir algún tipo de correlación.
2. Aplicar la prueba de independencia (con matrices particionadas).
3. Encontrar la correlación canónica entre los y_i 's y x_j 's y definir las variables canónicas.
4. Aplicar la prueba de significancia para las variables canónicas
5. Interpretar los coeficientes estandarizados para las variables canónicas
6. Interpretar las correlaciones entre las variables canónicas y las variables originales.
7. Estimar los valores de las variables canónicas en función a la muestra, posteriormente graficar estos resultados, para evidenciar la correlación entre las variables canónicas.

5. CUARTA FASE: Análisis de Correspondencias

5.1. Análisis de Correspondencias Simple

Construir la Tabla de contingencias entre dos variables que considere significativas para cierto tipo de análisis. Posteriormente desarrollar los siguientes puntos.

1. Realizar un análisis descriptivo analítico y gráfico de los datos (diagrama de barras y diagrama de mosaicos)
2. Aplicar la prueba de homogeneidad de filas o columnas (Prueba de independencia).
3. Graficar los datos en el espacio $k = 1, 2, 3$ dimensional adecuado, tomando en cuenta el porcentaje de inercia explicada en cada dimensión.
4. Identificar las categorías que tienen un aporte significativo en el estadístico de la χ^2 , tomando en cuenta las distancias chi-cuadrado de los perfiles fila y columna con respecto a su perfil medio.

5.2. Análisis de correspondencias Múltiple

Construir la Tabla de contingencias con tres variables que considerere significativas para cierto tipo de análisis, posteriormente desarrollar los siguientes puntos.

1. Aplicar la prueba de independencia e interpretar los resultados.
2. Utilizar la matriz binaria y graficar los datos en el espacio $k = 1, 2, 3$ dimensional adecuado, tomando en cuenta el porcentaje de inercia explicada en cada dimensión.
3. Utilizar la matriz de Burt y graficar los datos en el espacio $k = 1, 2, 3$ dimensional adecuado, tomando en cuenta el porcentaje de inercia explicada en cada dimensión.
4. Comparar los resultados obtenidos.