

**FACULTAD CIENCIAS FÍSICAS Y MATEMÁTICAS**

**INGENIERÍA EN INFORMÁTICA**

**ANÁLISIS DE DATOS**

**ANÁLISIS DEL DISCRIMINAN TE**

**GRUPO 5**

**INTEGRANTES:**

**GUAYASAMIN CINTHYA**

**HERRERA ESTEBAN**

**MIRANDA CHRISTIAN**

**ANÁLISIS DISCRIMINANTE .**

El Análisis Discriminante es una técnica estadística multivariante cuya finalidad es, analizar si existen diferencias significativas entre grupos de objetos respecto a un conjunto de variables medidas sobre los mismos, para en el caso de que existan, explicar en qué sentido se dan y facilitar procedimientos de clasificación sistemática de nuevas observaciones de origen desconocido en uno de los grupos analizados.

**OBJETIVOS**

•Se pretende encontrar relaciones lineales entre las variables continuas que mejor discriminen en los grupos dados a los objetos.

•Construir una regla de decisión que asigne un objeto nuevo con un cierto grado de riesgo, cuya clasificación previa se desconoce, a uno de los grupos prefijados.

Identificar si existen diferencias significativas entre grupos de objetos respecto a un conjunto de variables medidas sobre si mismos.

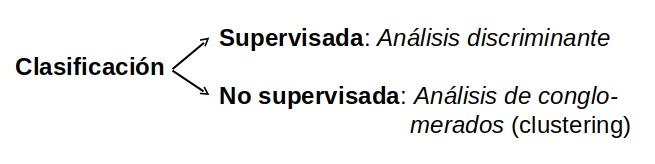
**APLICACIONES**

¿Se puede predecir si una empresa va a entrar en bancarrota?

¿Es posible predecir con antelación si un cliente que solicita un préstamo a un banco va a ser un cliente moroso?

¿Existe discriminación por razones de sexo o de raza en una empresa o en un colegio?

**INTRODUCCIÓN**

El análisis discriminante es una técnica de clasificación para asignar nuevas observaciones a grupos ya conocidos.

**VARIABLES CATEGÓRICAS E INDEPENDIENTES**

Las variables categóricas también se denominan variables cualitativas o variables atributos. Los valores de una variable categórica se pueden colocar en un número contable de categorías o grupos diferentes. Los datos categóricos pueden tener o no algún orden lógico.

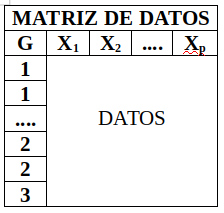
Los valores de una variable cuantitativa se pueden ordenar y medir.

**Supuestos**

* Existen K poblaciones o grupos G1, G2,....Gk.
* Cada grupo está formado por n1, n2,...nk objetos: .
* Sobre cada objeto han sido medidas p variables x1, x2,.xp,
* Se quiere buscar una regla de decisión que permita asignar un objeto a uno de los grupos partiendo de la información anterior.
* Los datos se presentan en matriz de n objetos pertenecientes a K grupos, medidos por una variable aleatoria p dimensional y una variable discreta que indica el grupo al que pertenece cada objeto.

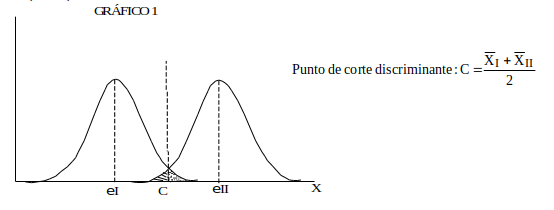
**Clasificación**

* **Funciones discriminantes lineales de Fisher**:  
  Permiten diferenciar los grupos para el proceso de clasificación. Son combinación lineal de las P variables, interviniendo cada una con un peso diferente que indica las que más discriminan. Problema descriptivo.
* **Funciones discriminantes** **canónicas:**  
  Sirven para la predicción óptima del grupo a que pertenece un individuo. Problema de inferencia.



**CLASIFICACIÓN CON DOS GRUPOS Y UNA VARIABLE CLASIFICADORA X**

Clasificar a cada individuo en el grupo correcto, según la variable clasificadora. Gráficamente, podríamos representar las hipotéticas funciones de frecuencias de la variable X para cada uno de los dos grupos. Las distribuciones de frecuencias y la varianza son iguales en los dos grupos, coincidiendo en todo menos en su media. Se pueden solapar: pueden haber errores de clasificación.



**CRITERIOS DE CLASIFICACIÓN**

Si Xi < C, se clasifica al individuo i en el grupo I.

Xi > C, se clasifica la individuo i en el grupo II

Los errores de clasificación se encuentran en:

Área a la derecha de C: Casos del grupo I en los que Xi > C, es decir, son casos del grupo I mal clasificados en el grupo II.

Área a la izquierda de C: Casos del grupo II en los que Xi < C, es decir, son casos del grupo II mal clasificados en el grupo I.

**CLASIFICACIÓN CON DOS O MAS GRUPOS Y DOS O MAS VARIABLES CLASIFICADORAS**

Para dos grupos, desde el punto de vista factorial el objetivo es obtener una función lineal que separe lo mejor posible a los dos grupos.

Criterio:

Buscar el eje que separe lo más posible los centros de los grupos, de forma que los individuos de cada grupo sean lo más homogéneos posibles. Hay que maximizar la dispersión entre grupos respecto a la dispersión dentro de los grupos.

Generalizar a K grupos:

Habrá más de un eje discriminante. El objetivo es representar a los n individuos de K grupos predefinidos, en un espacio de dimensión reducida (ejes discriminantes) de forma que los grupos proyectados en ese espacio estén bien diferenciados.

**OBTENCIÓN DE LAS FUNCIONES DISCRIMINANTES**

Criterio: Maximizar variabilidad entre grupos respecto a la de dentro de ellos.

Sea un eje definido por el vector unitario U, de dimensión p\*1. La proyección del individuo i-ésimo sobre él es el peso zi:

Proyección de los n individuos: Vector Z (nx1), (Z=XU), función discriminante de Fisher. Es una combinación lineal de las k variables explicativas originales.

Problema:Obtener los coeficientes de ponderación uj.

Hay que tener en cuenta que:

1. La matriz a diagonal izar no es simétrica: los vectores propios no son necesariamente ortogonales.

2. El número de variables o ejes discriminantes, F es como máximo el mínimo de [(K-1), p].

**MÉTODOS ITERATIVOS DE SELECCIÓN DE VARIABLES**

**Método de inclusión iterativa**

En cada paso se selecciona la variable que más contribuye a la separación de los grupos. El proceso se detiene si ninguna variable separa los grupos significativamente más de lo que ya estaban.

**Método de exclusión iterativa**

Se incluyen todas las variables y en cada paso se elimina la que menos contribuye a la separación de los grupos. El proceso se detiene cuando la exclusión de cualquiera de las variables hace disminuir significativamente la separación entre los grupos.

**Método mixto de inclusión-exclusión: Stepwise**

En cada etapa se evalúa la posibilidad de incluir una nueva variable como la de excluir alguna de las presentes según criterios prefijados. Su aplicación requiere definir previamente una regla de decisión, Landa de Wilks, cociente entre el determinante de la matriz de variación dentro de los grupos y el de la matriz de variación total.

**REQUISITOS PARA UN ANÁLISIS DISCRIMINANTE**

1. Tener formados los grupos a analizar a priori.

2. No debe existir multicolinealidad entre las variables, es decir, no debe existir combinación lineal entre las diferentes variables discriminantes.

3. El número máximo de funciones discriminantes que se pueden calcular no debe ser mayor al número de grupos menos 1, pudiendo ser igual número de variables discriminantes existentes.

4. Todas las variables discriminantes deben provenir de poblaciones con distribución normal.

5. Debe existir dispersión y variabilidad dentro de cada grupo. NO puede existir varianza igual a cero en ningún grupo.

6.El análisis discriminante es muy sensible a datos extremos, la falta de homocedasticidad se traduce en pruebas no confiables, especialmente cuando el tamaño de la muestra es muy pequeño.

**PASOS PARA UN ANÁLISIS DISCRIMINANTE**

**SELECCIÓN DE VARIABLES DISCRIMINANTES**

Puede hacerse a través de la identificación de variables importantes de modelos teóricos validados en investigaciones previas y, por otro lado, de la aplicación de pruebas estadísticas. Ambas vías son complementarias y no excluyentes. La selección estadística consiste en la combinación de un proceso de análisis descriptivo (cálculo de las medias y desviaciones típicas), correlación entre las variables y diferencias significativas entre los grupos. La correlación se comprobará dentro del grupo y entre grupos. Se utilizará λ de Wilks y la correlación canónica.

**SELECCIÓN DE LA FUNCIÓN DISCRIMINANTE**

La función discriminante no es otra cosa que un factor o una nueva variable, combinación lineal de las anteriores. Podremos obtener más de una función discriminante pero solo una permitirá hacer predicciones de las mejores clasificaciones futuras. Se utilizará el coeficiente de “auto valor” y los coeficientes estandarizados de las funciones discriminantes. La función discriminante no es otra cosa que un factor o una nueva variable, combinación lineal de las anteriores. Podremos obtener más de una función discriminante pero solo una permitirá hacer predicciones de las mejores clasificaciones futuras. Se utilizará el coeficiente de “auto valor” y los coeficientes estandarizados de las funciones discriminantes.

**CLASIFICACIÓN DE LOS SUJETOS U OBJETOS**

Para poder asignar un sujeto a un grupo u otro tenemos que diseñar una regla de decisión. Una vez aplicada se hará una valoración en porcentaje de aciertos en el pronóstico de clasificación de los sujetos y se obtendrá una tabla resumen de clasificación.

**OTRA METODOLOGÍA**

1. Seleccionar las variables independientes métricas y la variable dependiente no métrica a utilizar.

2. Seleccionar el tamaño muestral.

3. Dividir la muestra.

4. Obtener las constantes del modelo.

5. Probar el modelo

6. Evaluar la contribución de las variables.

7. Valorar de forma predictiva el modelo para verificar el grado de error que tendrá efectuar pronósticos.

8. Interpretar los resultados del modelo.

**PASOS PARA EL ANÁLISIS DISCRIMINANTE**

