

# Investigación de Mercados II

Lección 6: La clasificación de  
consumidores a través del análisis  
discriminante

# Contenido

1. Fundamentos del Análisis Discriminante
2. El Proceso del Análisis Discriminante
3. El Análisis Discriminante con SPSS.

# 1. Fundamentos del análisis discriminante

- El Análisis Discriminante (AD) encuentra funciones capaces de clasificar a individuos en grupos tomando como base un conjunto de medidas cuantitativas (variables escala) sobre los mismos.
- Dichas funciones, son combinaciones lineales de las variables independientes, y discriminarán o identificarán los grupos a través de la variable dependiente de dichas funciones.

# 1. Fundamentos del análisis discriminante

- Es un **método de dependencia**, por tanto **distingue entre variable dependiente (ha de ser categórica) y variables independientes (deben ser métricas)**.
- Los **grupos** que queden **definidos por la variable dependiente serán exhaustivos y mutuamente excluyentes**; es decir, cada uno de los casos pertenece única y exclusivamente a un grupo.

# 1. Fundamentos del análisis discriminante

*Similitudes y diferencias entre el ANOVA, la Regresión y el Análisis Discriminante*

<b>Variables</b>	<b>ANOVA</b>	<b>Regresión</b>	<b>Discriminante</b>
<b><i>Número</i></b>	<i>Similitudes</i>	<i>Similitudes</i>	<i>Similitudes</i>
<b>Dependientes</b>	Una	Una	Una
<b>Independientes</b>	Múltiples	Múltiples	Múltiples
<b><i>Naturaleza</i></b>	<i>Diferencias</i>	<i>Diferencias</i>	<i>Diferencias</i>
<b>Dependientes</b>	Cuantitativa	Cuantitativa	Categórica
<b>Independientes</b>	Categórica	Cuantitativa	Cuantitativa

# 1. Fundamentos del análisis discriminante

## Tipos de Análisis Discriminante:

Se clasifican en función al número de categorías de la variable dependiente en:

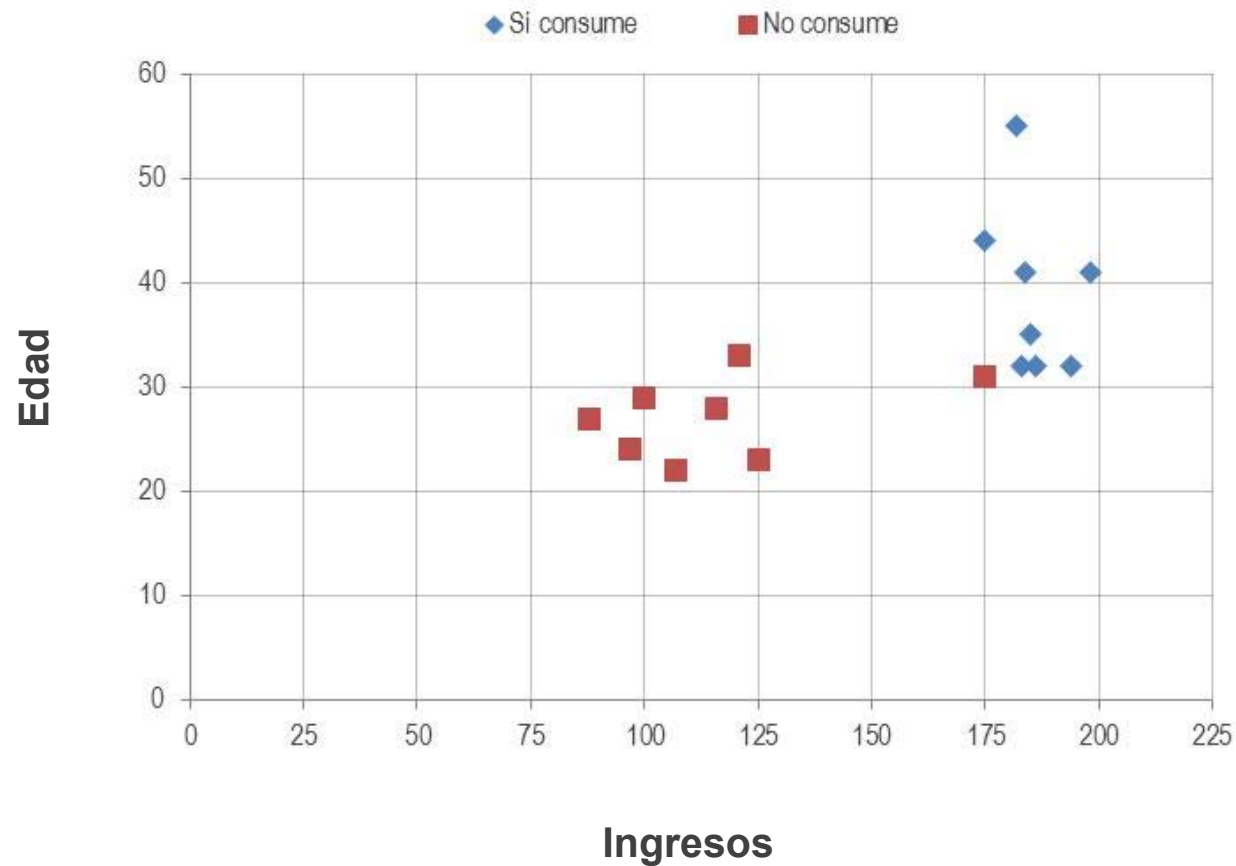
- 1. Análisis discriminante de dos grupos o simple:** la variable dependiente tiene sólo dos categorías (variables dummy, dicotómica o binaria).
  - En esta situación, se obtendrá solo una función discriminante
- 2. Análisis discriminante múltiple:** la variable dependiente tiene más de dos categorías.
  - Aquí, si fuese necesario, se pueden obtener más de una función discriminante, concretamente el mínimo entre  $[G-1, p]$ :
    - Siendo “G” el número de grupos de la variable dependiente.
    - Siendo “p” el número de variables independientes.

# 1. Fundamentos del análisis discriminante

**Ejemplo:** Una empresa tiene una base de datos de sus clientes con tres variables: consumo de su último producto lanzado al mercado, edad e ingresos. Desea encontrar una función que discrimine entre los consumidores y los no consumidores del producto que la misma comercializa.

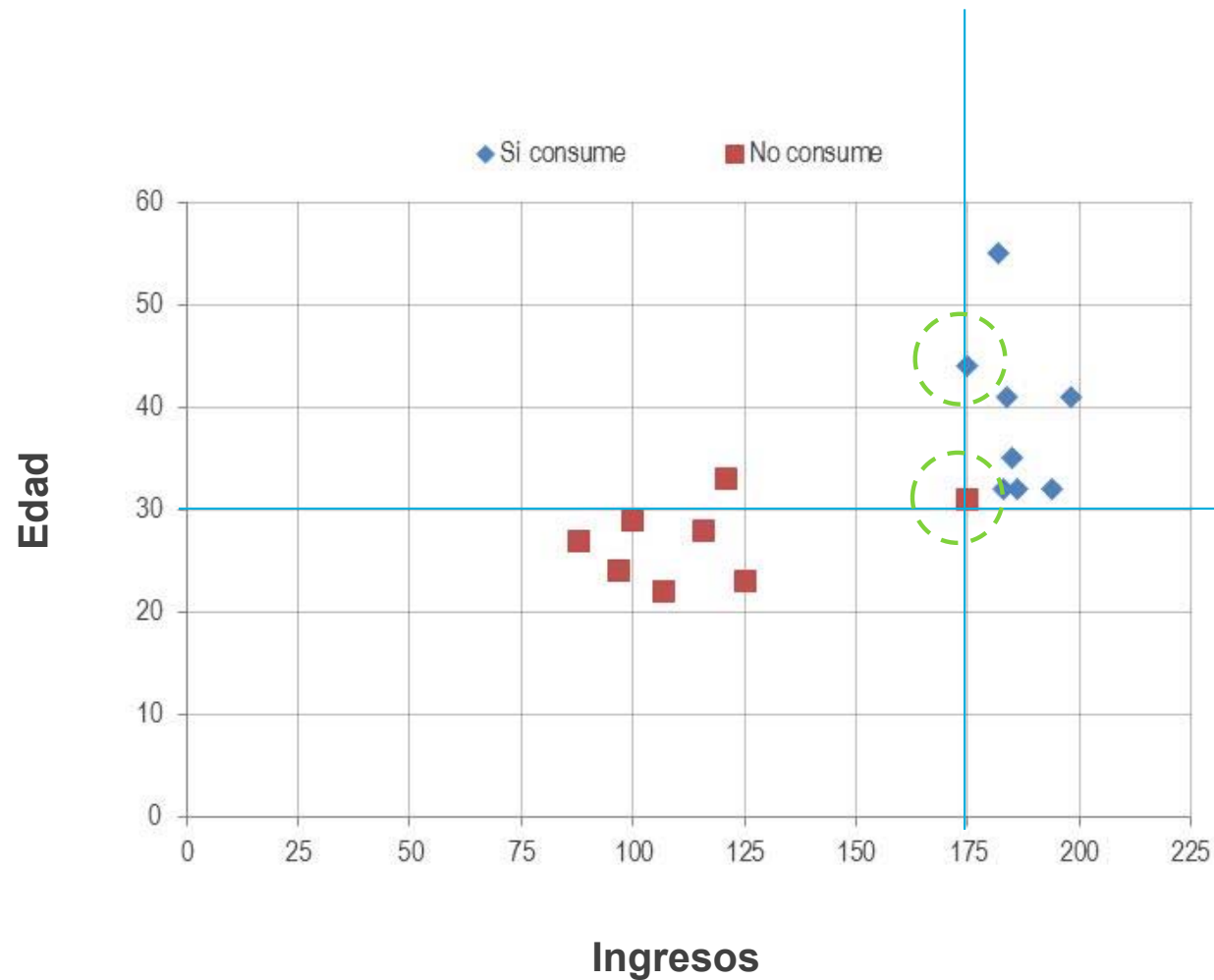
Individuo	Ingresos	Edad	Consume	Individuo	Ingresos	Edad	Consume
1	175	44	Sí	9	125	23	No
2	182	55	Sí	10	107	22	No
3	184	41	Sí	11	97	24	No
4	186	32	Sí	12	88	27	No
5	185	35	Sí	13	116	28	No
6	198	41	Sí	14	121	33	No
7	194	32	Sí	15	175	31	No
8	183	32	Sí	16	100	29	No

# 1. Fundamentos del análisis discriminante





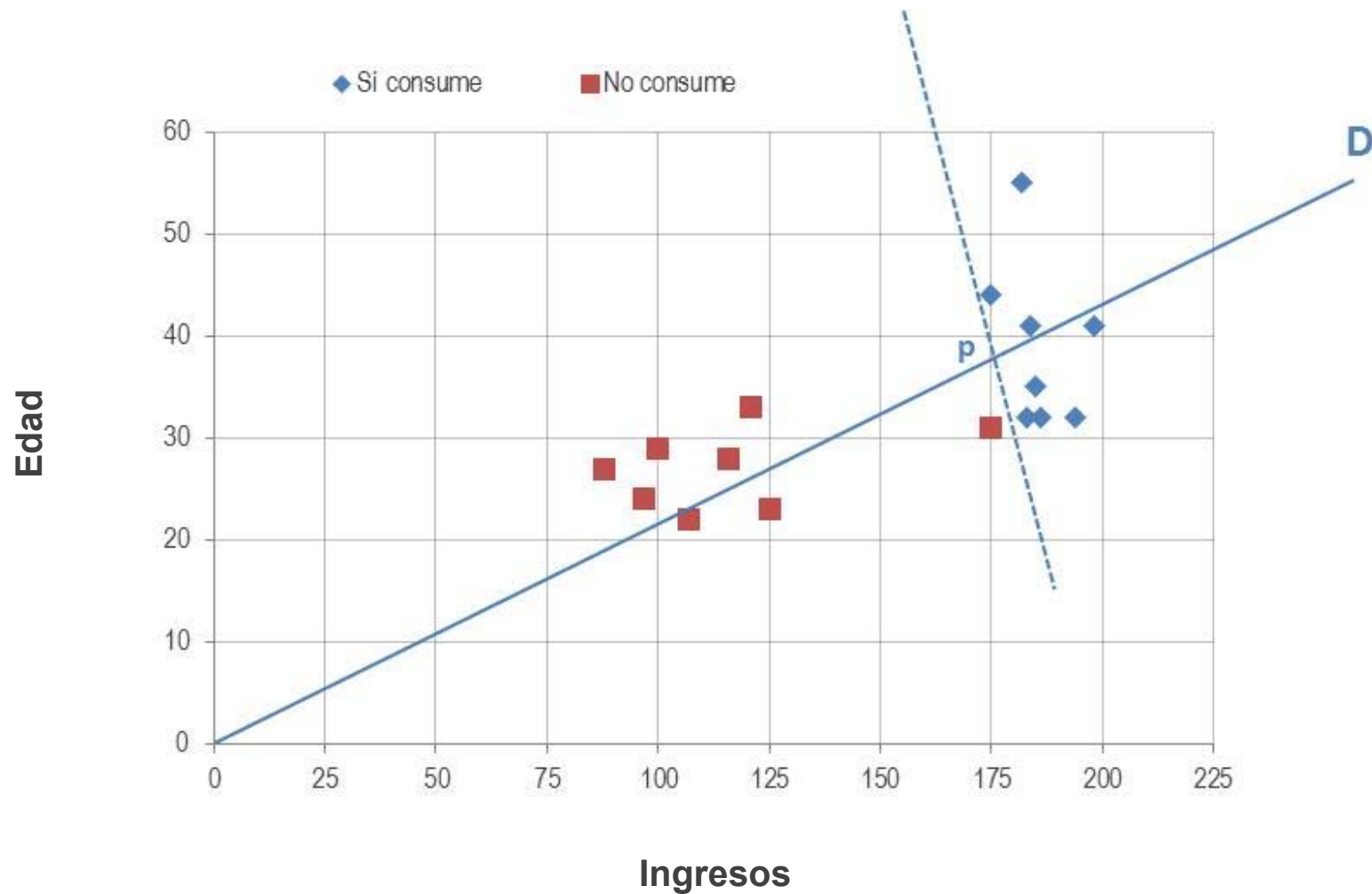
# 1. Fundamentos del análisis discriminante



# 1. Fundamentos del análisis discriminante

- En general, **las dos variables son buenas para discriminar** entre consumidores y no consumidores del producto de la empresa.
- **Podríamos concluir que clientes con ingresos superiores a 175 y de más de 30 años son los consumidores del producto. Pero esas conclusiones conllevan a errores de clasificación.**
- Lo ideal sería **encontrar una combinación lineal de las variables ingresos y edad, que discrimine mejor entre ambos grupos.**

# 1. Fundamentos del análisis discriminante



# 1. Fundamentos del análisis discriminante

## Utilidades del Análisis Discriminante:

- 1. Explicativa:** cuantifica la contribución relativa de cada una de las variables independientes en la clasificación correcta de los individuos.
- 2. Predictiva:** encasilla a un individuo dentro de un grupo, a partir de los valores de las variables independientes.
- 3. Re-clasificadora:** definidos grupos a priori (por ejemplo a través de clústers), se desea verificar dicha partición, corroborándola o corrigiéndola a través del análisis discriminante.

# 1. Fundamentos del análisis discriminante

## Supuestos de Partida del Análisis Discriminante:

1. **Matrices de Covarianzas Intra-Grupos Iguales** en los grupos definidos por la variable dependiente.
  - Esto es necesario, ya que para la clasificación de los casos utiliza **la matriz de covarianzas combinada de los grupos**.
  - El mayor problema de que no se cumpla este supuesto es que el modelo tenderá a **clasificar casos en los grupos con mayor matriz de covarianzas**.
  - Para comprobar este supuesto se suele recurrir al **test M de Box**. De forma intuitiva, se puede observar la similitud de **los logaritmos de los determinantes** de dichas matrices.
  - **Si el supuesto no se cumple**, podemos modificar el método utilizando para la **clasificación matrices de covarianzas de los grupos por separado**.

# 1. Fundamentos del análisis discriminante

## Supuestos de Partida del Análisis Discriminante:

2. **Normalidad Multivariante:** cada grupo ha de proceder de una muestra de una población que siga una **distribución normal multivariante**.
  - Para comprobar la hipótesis de normalidad multivariante, **se examinan si las variables independientes siguen individualmente una distribución normal**. De ser así, conjuntamente se distribuirán como una normal multivariante.
  - **En caso de que no se cumpla, provocará problemas** en la interpretación de las funciones discriminantes, sobre todo porque **los tests de significación no serán validos**. En estos casos, se recomienda **utilizar una regresión logística**.

# 1. Fundamentos del análisis discriminante

## Supuestos de Partida del Análisis Discriminante:

3. **Ausencia de Multicolinealidad:** las variables independientes no están altamente correlacionadas.
  - En caso de que **exista multicolinealidad**, algunas variables independientes **podrían estar explicadas por otras**, lo que afectaría a su capacidad predictiva.
  - Para chequear esto, **se utiliza la matriz de correlaciones** y a posteriori, ver si se producen diferencias entre **la matriz de estructura de correlaciones** (no alterada por la multicolinealidad) con respecto a los coeficientes estandarizados de la función discriminante.

# 1. Fundamentos del análisis discriminante

## Recomendaciones respecto a la muestra:

- Estadísticamente representativa:
  - a) Mínimo de 20 observaciones para cada grupo y para cada variable independiente.
  - b) El grupo más pequeño en miembros debería exceder al número de variables independientes.
- Un caso será excluido del análisis si no se tiene información sobre el mismo acerca de las variables que definen el modelo.



## 2. El Proceso del Análisis Discriminante

### 1. Selección del método:

**a) Método directo o simultáneo: todas las variables independientes son incluidas simultáneamente.**

- Se suele utilizar cuando por razones teóricas se quiere que todas las variables intervengan en el análisis.
- También cuando se desconoce o no interesa el poder discriminatorio de cada una de ellas.

## 2. El Proceso del Análisis Discriminante

### 1. Selección del método:

b) **Método por pasos:** consiste en **retener el mejor conjunto de variables independientes que discriminen entre grupos.**

- El método **comienza eligiendo la variable que mejor discrimina** entre grupos individualmente, y secuencialmente, va incluyendo otras que conjuntamente con las ya seleccionadas discriminen mejor.
- Existen muchos criterios de inclusión de variables (distancia de Mahalanobis, Menor Razón F, V de Rao), aunque el más utilizado **es el que minimiza la Lambda de Wilks:**

## 2. El Proceso del Análisis Discriminante

### 1. Selección del método:

- Método por pasos minimizando la Lambda de Wilks:

$$MIN (\Lambda) = \left( \frac{SCI}{SCT} \right) = \left( \frac{SCI}{SCE + SCI} \right)$$

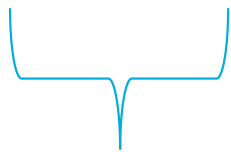
- Es decir, selecciona la variable o variables que consigan **grupos muy separados entre sí** (maximizar la Suma de Cuadrados Entre Grupos) e **internamente homogéneos** (minimizar la Suma de Cuadrados Intra Grupos).
- El criterio para la inclusión o eliminación de las variables en el modelo es el de **Tolerancia**, que muestra **el porcentaje de la variable no explicado por el resto de las variables del modelo**.
  - Si Tolerancia = 1 (variable totalmente independiente).
  - Si Tolerancia= 0 (variables totalmente dependiente).

## 2. El Proceso del Análisis Discriminante

### 1. Selección del método:

- Independientemente del método de inclusión de variables, la función discriminante será de esta manera:

$$D = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_p X_p$$



Variable categórica



Variables métricas o cuantitativas

## 2. El Proceso del Análisis Discriminante

### 2. Evaluación de la Significación de las Funciones:

- Consiste en **determinar si la función o funciones discriminantes son significativas**, es decir, si realmente son capaces de discriminar los grupos de individuos. Para ello, se utilizan:
  - a) Los **autovalores de la matriz de covarianzas de la función discriminante** que muestran la **eficacia relativa de cada función discriminante**.
  - b) Las **correlaciones canónicas**: equivalente a la **correlación de Pearson entre la puntuación en la función discriminante y la pertenencia a un grupo**.

## 2. El Proceso del Análisis Discriminante

### 2. Evaluación de la Significación de las Funciones:

- c) La **Lambda de Wilks** de la función explica el **porcentaje de la varianza total de los datos que no es explicada por la diferencia entre grupos**, por tanto:
- **Si es próximo a cero**, las funciones D son capaces de reflejar mucha diferencia entre-grupos y poca intra-grupos, es decir **separan muy bien los grupos**.
  - **Si es próxima a uno**, las funciones D **no reflejan diferencias entre grupos**.

## 2. El Proceso del Análisis Discriminante

### 2. Evaluación de la Significación de las Funciones:

- c) La **Lambda de Wilks de la función** se puede aproximar por una Chi- Cuadrado que testea la hipótesis de igualdad de medias de puntuación entre grupos.
  - Por tanto, si No Rechazo la  $H_0$ , las medias de puntuación entre los grupos serán iguales y la función no discriminará bien.

## 2. El Proceso del Análisis Discriminante

3. **Estudio de las Funciones Discriminantes:** si la función discriminante, entonces podremos:
  - a) **Calcular el Punto de Corte de la función:** determina el valor a partir de cual un caso va a un grupo u otro. Se utiliza si existen **solo dos grupos** en la variable dependiente.
    - Si el tamaño de los grupos es igual, es la media aritmética de los centroides de ambos grupos.
    - Si el tamaño de los grupos es distinto, es la media ponderada de los centroides de ambos grupos.



## 2. El Proceso del Análisis Discriminante

### 3. Estudio de las Funciones Discriminantes:

#### b) Ajustar las Probabilidades A Priori:

- Por defecto, se presupone que **existe la misma probabilidad de pertenecer a un grupo** u otro de la variable dependiente.
- Aunque, **en otras ocasiones, el tamaño del grupo afecta a la clasificación de los casos**. El investigador puede optar por considerar **probabilidades en función al tamaño de los grupos**.

## 2. El Proceso del Análisis Discriminante

### 3. Estudio de las Funciones Discriminantes:

#### c) Estudiar las funciones lineales de Fischer:

- Se calcula una para cada grupo y ayudan a la clasificación de los casos.
- Un caso irá al grupo al que le pertenezca la función de Fischer donde más puntúe.

## 2. El Proceso del Análisis Discriminante

### 3. Estudio de las Funciones Discriminantes:

#### d) Visualizar el Mapa Territorial:

- Se trata de un gráfico que representa las funciones discriminantes (en caso de que haya más de una) en los ejes.
- En él quedan reflejados los centroides de los grupos, así como las fronteras que sirven para separar entre ellos.

## 2. El Proceso del Análisis Discriminante

4. **Idoneidad de la Clasificación:** se trata de determinar el total de casos bien clasificados por la función. Para ello se recurre a **la matriz de confusión o tabla de clasificación**, que muestra **los casos bien clasificados (en la diagonal principal)** y mal clasificados (fuera de la diagonal principal).
  - El ratio de casos bien clasificados recibe el nombre de **“Hit Ratio”** y se utiliza como **indicador de la capacidad discriminante de la función** (similar al R cuadrado).
  - Como el Hit Ratio suele estar **sesgado al alza**, además se utiliza la **clasificación dejando uno fuera**, que clasifica cada caso del análisis mediante la función derivada de todos los casos, excepto el propio caso. También se conoce como **método U**.

## 2. El Proceso del Análisis Discriminante

### 5. Interpretación de la función discriminante:

- **Coeficientes estandarizados de la función discriminante:** permite comparar coeficientes de variables medidas en diferentes magnitudes. Las variables con coeficientes estandarizados mayores contribuirán más a la capacidad de discriminación de la función.
- **Matriz de estructura de correlaciones:** muestra la correlación entre la función discriminante y una variable independiente. A mayor correlación, más importante el papel de la variable en la función discriminante.
- **ANOVAS Univariados:** valores pequeños de la Lambda implicarán un valor elevado de F, que significará que la variable tiene un alto poder discriminante. Por otro lado, si la Sig. > 5%, la variable no discrimina bien, se puede considerar su eliminación del análisis.