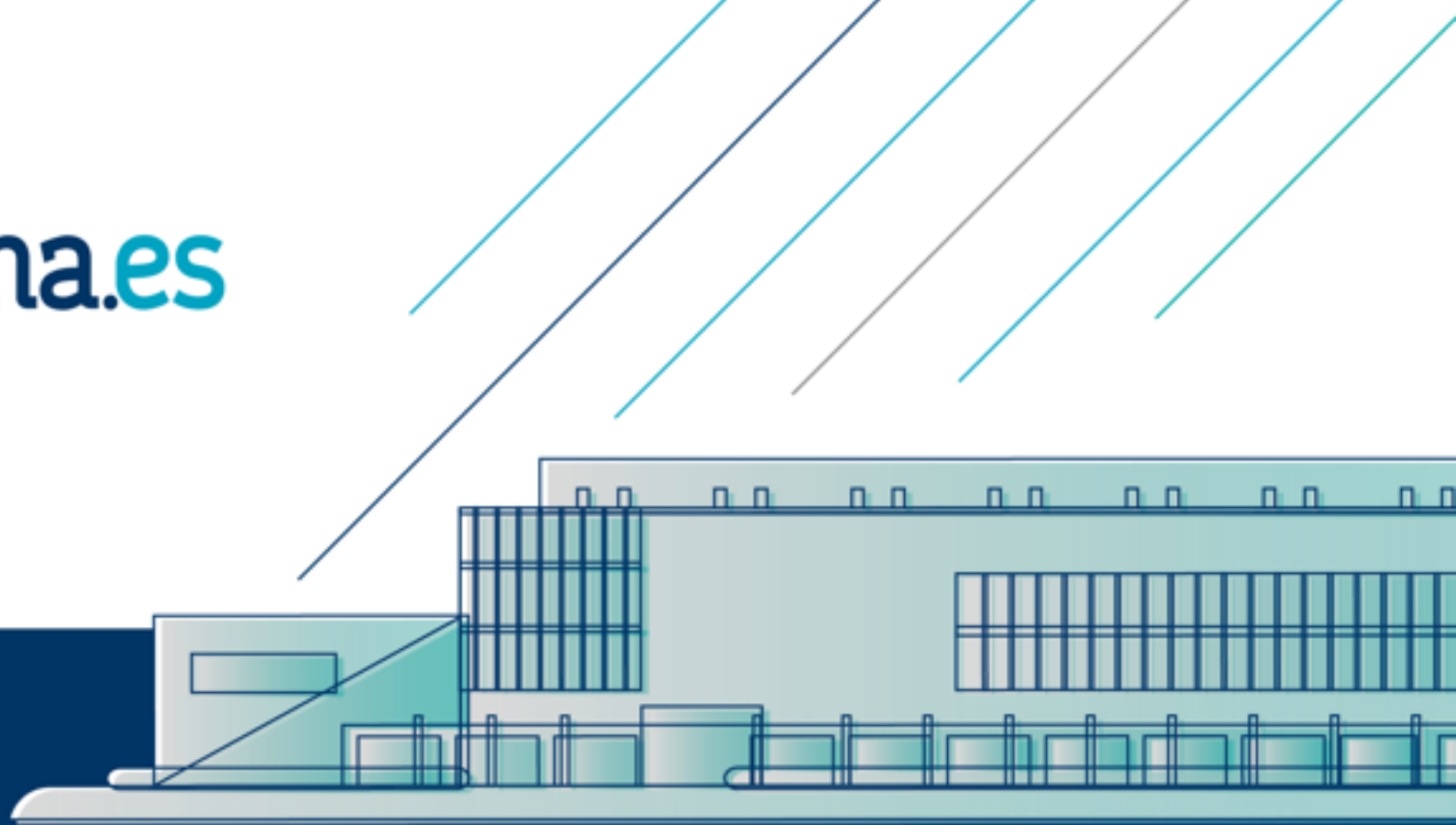




UNIVERSIDAD  
DE MÁLAGA

| [uma.es](http://uma.es)



## Tema 5 (Parte 2): La clasificación de consumidores a través de la regresión logística

# ¿Por qué necesitamos la regresión logística?



En investigación de mercados y análisis de clientes es habitual que el resultado de interés sea binario: comprar / no comprar, impagar / no impagar, abandonar / no abandonar.



En estos casos interesa **estimar la probabilidad** de que ocurra un suceso, y analizar **cómo influyen distintas variables explicativas** en dicha probabilidad.



La regresión logística binaria permite abordar este tipo de problemas de forma flexible y robusta.

# El problema de partida

---



Disponemos de individuos para los que: la **variable dependiente es binaria** (0/1), y contamos con **variables explicativas métricas y/o categóricas**.



El objetivo es: **modelizar la probabilidad** de que ocurra el suceso de interés, y utilizar dicha probabilidad para **clasificar y tomar decisiones**.

# ¿Qué hace exactamente la Regresión Logística Binaria?

---

01

Modeliza la probabilidad de que  $Y = 1$

02

Garantiza que las probabilidades estén acotadas entre 0 y 1.

03

Permite analizar el efecto individual de cada variable explicativa.

04

Proporciona un modelo interpretable para la toma de decisiones.

- +
  - 
  -
- A diferencia de otros métodos de clasificación, no requiere supuestos estrictos de normalidad.



# Regresión logística versus Análisis discriminante

- La regresión logística es una alternativa natural al análisis discriminante cuando los supuestos no se cumplen.

Análisis Discriminante	Regresión Logística
Clasifica individuos	Estima probabilidades
Requiere supuestos fuertes	No exige normalidad
Mejor con variables métricas	Acepta métricas y categóricas
Sensible a la M de Box	Más robusta

# Las 3 fases de la regresión logística

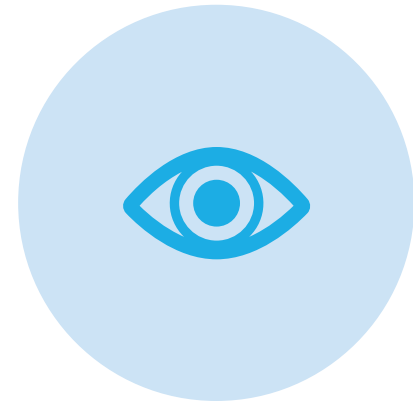
---



ESPECIFICACIÓN Y VALIDACIÓN  
DEL MÉTODO



EVALUACIÓN DE LA  
SIGNIFICACIÓN Y DEL AJUSTE



INTERPRETACIÓN DE  
RESULTADOS Y CLASIFICACIÓN

# FASE 1: ESPECIFICACIÓN DEL MODELO

---

Para aplicar la regresión logística binaria es necesario que:

La variable dependiente sea binaria.

---

Las variables independientes estén justificadas teóricamente.

---

Las variables categóricas se recodifiquen en variables ficticias.

---

El modelo tenga sentido desde el punto de vista del problema analizado.

---

# FASE 1: VARIABLES CATEGÓRICAS Y VARIABLES FICTICIAS

Cuando una variable categórica tiene  $k$  categorías, es necesario:

Definir  $k-1$  variables ficticias.

Seleccionar una categoría de referencia.

Interpretar los coeficientes en comparación con dicha referencia.





## Fase 2: Evaluación del ajuste del modelo



Para evaluar si el modelo es adecuado se utilizan:



La prueba de Hosmer-Lemeshow

No significativa = Buen ajuste del modelo.



Medidas de pseudo- $R^2$ : indican la capacidad explicativa del modelo.

$R^2$  de Cox y Snell.  
 $R^2$  de Nagelkerke.

# Fase 3: Interpretación de resultados

---

La interpretación del modelo se basa en:

El estadístico de Wald, que indica la significación individual de cada variable.

---

El odds ratio ( $\text{Exp}(B)$ ), que mide el efecto de cada variable sobre la probabilidad del suceso.

---

La tabla de clasificación, que resume la capacidad predictiva del modelo.

---

# Interpretación de odds ratio

---



El odds ratio indica cuánto se incrementa o reduce la probabilidad del suceso ante un cambio unitario en la variable explicativa, manteniendo el resto de las variables del modelo.



Su interpretación debe hacerse siempre en términos comparativos.

# Probabilidad frente a clasificación

La regresión logística no solo clasifica individuos: dos individuos pueden clasificarse en el mismo grupo, pero tener probabilidades muy distintas de que ocurra el suceso.



Este enfoque probabilístico aporta un valor añadido clave frente a otros métodos de clasificación.



# Aplicación práctica con SPSS

Regresión logística

Dependientes:

Bloque 1 de 1

Covariables:

VGB

Método: Intro

Variable de selección:

Regresión logística: Definir variables categóricas

Covariables:

Covariables categóricas:

VGB(Indicador(primera))

Cambiar el contraste

Contraste: Indicador

Categoría de referencia:  Primero  Último

Contínuar Cancelar Ayuda

Variable	Tipo
Escala	Entrada
Escala	Entrada
Escala	Entrada
Escala	Entrada
Escala	Entrada
Escala	Entrada
Escala	Entrada
Escala	Entrada
Escala	Entrada
Escala	Entrada
Escala	Entrada
Escala	Entrada

Variable	Tipo
Me permite desahogarme el derrotar a otros usuarios.	{1, T
Me gusta demostrar en un videojuego todo lo que soy capaz ...	{1, T
Jugar a videojuegos me permite liberar tensiones.	{1, T
Jugar a videojuegos se convierte en un modo de desconexió...	{1, T

# Aplicación práctica con SPSS

Regresión logística

Dependientes:

Usuario ID [ID]  
Gameplay genre breadth [G...  
Viewing genre breadth [VGB]  
Device use breadth [DUB]

Bloque 1 de 1

Anterior

Siguiente

Categorica...  
Guardar...  
Opciones...  
Estilo...

Regresión logística: Opciones

Estadísticos y gráficos

Gráficos de clasificación

Bondad de ajuste de Hosmer-Lemeshow

Listado de residuales por caso

Correlaciones de estimaciones

Historial de iteraciones

CI para exp(B): 95 %

Valores atípicos fuera 2 desviación estándar

Todos los casos

Mostrar

En cada paso  En el último paso

Probabilidad para el método por pasos

Entrada: 0,05 Eliminación: 0,10

Punto de corte para la clasificación: 0,5

Iteraciones máximas: 20

Conservar memoria para análisis complejos o conjuntos de datos grandes

Incluir constante en modelo

Continuar Cancelar Ayuda

# Ejercicio 1: Probabilidad de compra en ecommerce (probabilidad\_compra.sav)

- Un comercio electrónico desea analizar qué factores influyen en la probabilidad de que un cliente realice una compra. Para ello, dispone de información sobre el comportamiento de sus usuarios, así como de una variable que indica si el cliente ha comprado o no. Con la base de datos proporcionada:
  - Justifique el uso de un modelo de análisis multivariante.
  - Estime y analice los resultados del modelo propuesto.
  - Concluya sobre qué factores son los que más influyen en la probabilidad de compra del cliente.

## Ejercicio 2: Probabilidad de impago de clientes de entidad financiera (probabilidad\_impago.sav)

- Una entidad financiera desea identificar qué factores influyen en la probabilidad de que un cliente incurra en impago en los préstamos concedidos. Para ello, dispone de información socioeconómica que indica si el cliente ha tenido o no impago.
- Con la información proporcionada:
  - Justifique el uso de un modelo de análisis multivariante.
  - Estime y analice la capacidad de ajuste del modelo propuesto.
  - Concluya sobre qué perfil de cliente presenta una mayor probabilidad de impago.

# Ejercicio 3: Probabilidad de abandono de clientes (probabilidad\_abandono.sav)

- Una empresa de servicios digitales quiere analizar los factores que influyen en la probabilidad de abandono de sus clientes. Para ello dispone de información sobre el uso del servicio, el nivel de satisfacción, la edad del cliente y el canal principal de relación con la empresa. Con el objetivo de construir un modelo que permita estimar la probabilidad de abandono y apoyar la toma de decisiones comerciales:
  - Justifique el uso de un modelo de análisis multivariante para abordar el problema planteado.
  - Estime y analice la significación del modelo.
  - Interprete los resultados y extraiga conclusiones relevantes para la gestión de clientes.



UNIVERSIDAD  
DE MÁLAGA

| [uma.es](http://uma.es)