

Departamento de Estadística e I.O.

Máster en Estadística Aplicada



**UNIVERSIDAD
DE GRANADA**

**MODELOS DE RESPUESTA DISCRETA
APLICACIONES BIOSANITARIAS**

Tema 1

Guía de trabajo autónomo

Profesores

Ana María Aguilera del Pino

Manuel Escabias Machuca

Título original: Modelos de Respuesta Discreta. Aplicaciones Biosanitarias.
Guía de Trabajo Autónomo del Tema 1

© Los profesores

Todos los derechos reservados. Esta publicación es de uso personal del alumno y no puede ser reproducida, ni registrada, ni transmitida en ninguna forma ni por ningún medio, sin el permiso de los autores

Índice general

1. Guía de Trabajo Autónomo del Tema 1	1
1.1. Justificación del Tema 1	1
1.2. Objetivos del Tema 1	2
1.3. Actividades del Tema 1	2

Capítulo 1

Guía de Trabajo Autónomo del Tema 1

Para asimilar los contenidos del Tema 1 el alumno debe estudiar los apuntes de teoría del tema y realizar las actividades de trabajo autónomo de esta guía. Estas actividades tienen como objetivo facilitar el autoaprendizaje del alumno y la consecución de los objetivos del tema. Es conveniente que el alumno las trabaje individualmente y plantee en el foro sus posibles dudas e inquietudes relacionadas con las mismas.

1.1. Justificación del Tema 1

En este tema se comienza por el caso más simple de respuesta discreta, planteándose el problema de predecir una variable de respuesta binaria a partir de una o varias variables relacionadas. Dado que la media de este tipo de respuestas es la probabilidad de que ocurra el suceso de interés, el modelo de regresión lineal es inviable para resolver este problema. Los modelos de respuesta binaria más usuales serán revisados en este tema, situándolos además en el contexto general de los modelos lineales generalizados.

Los modelos de respuesta binaria son de una gran aplicación en el campo de la epidemiología donde es usual estar interesado en estudiar el padecimiento de una enfermedad en función de un factor de riesgo, controlando a su vez otras variables relacionadas que pueden confundir o modificar el efecto del factor de riesgo sobre la enfermedad. En relación a la primera fase de un estudio epidemiológico se estudiarán los diseños muestrales usuales.

1.2. Objetivos del Tema 1

- Conocer distintos modelos estadísticos de estimación de una variable de respuesta binaria: modelos logit, probit y de valores extremos.
- Situar los modelos de respuesta binaria en el contexto más general de los modelos lineales generalizados.
- Estudiar las tres fases de un estudio epidemiológico y analizar sus características principales.
- Identificar los diseños muestrales usuales en las aplicaciones biosanitarias: prospectivos, transversales y retrospectivos (casos y controles).
- Distinguir entre el riesgo relativo de ocurrencia de un suceso y el cociente de ventajas a favor de dicho suceso.

1.3. Actividades del Tema 1

1. Haga una lectura comprensiva de los apuntes del Capítulo 1 para obtener una visión global del tema.
2. Haga un resumen de las razones por las que el modelo de regresión lineal no es válido para explicar una respuesta binaria a partir de una variable relacionada (apartado 1.2).
3. Piense un ejemplo real de una variable de respuesta binaria (Y) y una variable explicativa cuantitativa (X) en la que se perciba claramente que variaciones iguales en X no producen variaciones iguales en la probabilidad de respuesta $Y=1$.
4. Interprete con sus propias palabras la relación que hay entre una respuesta binaria y un predictor cuantitativo en función del signo del parámetro beta asociado a la pendiente de la parte lineal del modelo. ¿Qué ocurre cuando beta es nulo?
5. ¿Cuál cree que es la diferencia fundamental entre los modelos logit y los probit? ¿En qué casos son aplicables los modelos de valores extremos? (apartado 1.3).
6. Piense en la razón por la que la función ligadura (link) de un modelo de respuesta discreta es la inversa de la función de distribución de una variable aleatoria continua.

7. Vea los modelos de respuesta discreta como casos particulares de los modelos lineales generalizados (apartado 1.4).
8. En experimentos toxicológicos en los que la probabilidad de morir es modelizada como una función de la dosis, el valor de la dosis x para el que la probabilidad de morir es $1/2$, $p(x) = 1/2$, es llamada dosis mediana letal y denotada por $LD50$.
 - a) Demostrar que tanto para el modelo logit como para el modelo probit $LD50 = -\alpha/\beta$.
 - b) Encontrar $LD50$ para el modelo de valores extremos con ligadura log-log.
9. Supongamos que se verifica el modelo con ligadura log-log (1.7). Explicar como se interpretaría el parámetro β .
10. Lea atentamente el apartado 1.5 y distinga con claridad las tres fases de un estudio epidemiológico y los tres tipos de diseño muestral que se pueden utilizar para generar los datos.
11. Conteste razonadamente la siguiente pregunta: ¿por qué en los estudios retrospectivos no tiene sentido estimar el riesgo relativo de ocurrencia de un suceso? (apartado 1.6)
12. Consideremos un estudio que pretende investigar la relación entre el infarto de miocardio y la condición de fumador. Para ello se ha realizado un estudio de casos y controles, donde los casos son 262 mujeres de mediana edad (< 69 años) ingresadas en unidades de cuidados intensivos en el norte de Italia con infarto de miocardio agudo durante 1983-1988. Cada caso lleva asociados dos controles ingresados en el mismo hospital con otras enfermedades agudas. Todos estos individuos fueron clasificados de acuerdo a si habían fumado alguna vez en la siguiente tabla:

Fumador	Infarto	
	Si	No
Si	172	173
No	90	346

Estimar aproximadamente el riesgo relativo de infarto en base al cociente de ventajas asociado, y teniendo en cuenta la imposibilidad de una estimación directa por tratarse de un estudio retrospectivo.

13. Con la intención de explicar la probabilidad de morir en accidente de tráfico en base al uso o no de cinturón de seguridad, se dispone de la siguiente tabla referente a 577.006 accidentes de tráfico:

Cinturón	Muerte en accidente	
	Si	No
No	1601	162.527
Si	510	412.368

Demostrar que el riesgo relativo de muerte en accidente de tráfico es aproximadamente ocho veces mayor para los que no usaron cinturón que para los que si lo hicieron.

14. Lea el fichero 'IntroducciónR.pdf' que contiene un pequeño manual para introducirse en el software estadístico R que será utilizado en las prácticas con ordenador de la asignatura. Al final del fichero se proponen una serie de ejercicios que el alumno debe realizar para familiarizarse con el manejo de las operaciones básicas y las estructuras de datos de este lenguaje de programación.