

Análisis de regresión logística

Albert Sorribas

Grup de Bioestadística i Biomatemàtica

Facultat de Medicina, UdL



Análisis de regresión logística

- ☐ Concepto de riesgo relativo
- ☐ Odds ratio y riesgo relativo
- ☐ Necesidad del planteamiento multivariante
- ☐ Modelo de regresión logística
 - Definición
 - Estimación del riesgo relativo
 - Interpretación de resultados
- ☐ ¿Cómo realizar un análisis de regresión logística en SPSS?

El concepto de riesgo relativo

Tabla de contingencia GRUPO * ENFERMO

			ENFERMO		Total
			Si	No	
GRUPO	Expuestos	Recuento	23	12	35
		% de GRUPO	65,7%	34,3%	100,0%
	No expuestos	Recuento	15	26	41
		% de GRUPO	36,6%	63,4%	100,0%
Total		Recuento	38	38	76
		% de GRUPO	50,0%	50,0%	100,0%

$$\left. \begin{array}{l} P(S / E) = 0.657 \\ P(S / \bar{E}) = 0.366 \end{array} \right\} \rightarrow RR = \frac{P(S / E)}{P(S / \bar{E})} = 1.795$$

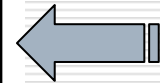
Ejemplo

Tabla de contingencia EXPOSICI * ENFERMO

			ENFERMO		Total
			Enfermos	No enfermos	
EXPOSICI	Expuestos	Recuento	45	22	67
		% de EXPOSICI	67,2%	32,8%	100,0%
	No expuestos	Recuento	25	56	81
		% de EXPOSICI	30,9%	69,1%	100,0%
Total		Recuento	70	78	148
		% de EXPOSICI	47,3%	52,7%	100,0%

Estimación de riesgo

	Valor	Intervalo de confianza al 95%	
		Inferior	Superior
Razón de las ventajas para EXPOSICI (Expuestos / No expuestos)	4,582	2,288	9,175
Para la cohorte ENFERMO = Enfermos	2,176	1,509	3,139
Para la cohorte ENFERMO = No enfermos	,475	,327	,689
N de casos válidos	148		



- ❑ El IC del riesgo relativo es (1.51 , 3,14)
- ❑ Podemos concluir que la exposición aumenta entre 1.51 y 3.14 veces la probabilidad de padecer la enfermedad
- ❑ La interpretación **médica** depende del problema (¿Hasta qué punto este aumento en la probabilidad es importante?)

Ejemplo

La interpretación depende de cómo se han codificado las variables

Tabla de contingencia EXPO * STATUS

		STATUS		Total
		1	2	
EXPO 1	Recuento	23	12	35
	% de EXPO	65,7%	34,3%	100,0%
2	Recuento	15	33	48
	% de EXPO	31,3%	68,8%	100,0%
Total	Recuento	38	45	83
	% de EXPO	45,8%	54,2%	100,0%

Tabla de contingencia EXPO * STATUS

		STATUS		Total
		0	1	
EXPO 0	Recuento	33	15	48
	% de EXPO	68,8%	31,3%	100,0%
1	Recuento	12	23	35
	% de EXPO	34,3%	65,7%	100,0%
Total	Recuento	45	38	83
	% de EXPO	54,2%	45,8%	100,0%

Estimación de riesgo

	Valor	Intervalo de confianza al 95%	
		Inferior	Superior
Razón de las ventajas para EXPO (1 / 2)	4,217	1,668	10,657
Para la cohorte STATUS = 1	2,103	1,297	3,409
Para la cohorte STATUS = 2	,499	,303	,820
N de casos válidos	83		

Estimación de riesgo

	Valor	Intervalo de confianza al 95%	
		Inferior	Superior
Razón de las ventajas para EXPO (0 / 1)	4,217	1,668	10,657
Para la cohorte STATUS = 0	2,005	1,220	3,295
Para la cohorte STATUS = 1	,476	,293	,771
N de casos válidos	83		

$$\frac{1}{2.103} = 0.476$$

$$\frac{1}{0.499} = 2.005$$

En la práctica, la frecuencia de la patología suele ser baja

Tabla de contingencia EXPOSICI * ENFERMO

			ENFERMO		Total
			Enfermos	No enfermos	
EXPOSICI	Expuestos	Recuento	7	134	141
		% de EXPOSICI	5,0%	95,0%	100,0%
	No expuestos	Recuento	2	154	156
		% de EXPOSICI	1,3%	98,7%	100,0%
Total		Recuento	9	288	297
		% de EXPOSICI	3,0%	97,0%	100,0%

Estimación de riesgo

	Valor	Intervalo de confianza al 95%	
		Inferior	Superior
Razón de las ventajas para EXPOSICI (Expuestos / No expuestos)	4,022	,822	19,694
Para la cohorte ENFERMO = Enfermos	3,872	,818	18,334
Para la cohorte ENFERMO = No enfermos	,963	,923	1,004
N de casos válidos	297		

□ Es necesario disponer de muchas observaciones

En la práctica, la frecuencia de la patología suele ser baja

Tabla de contingencia EXPOSICI * ENFERMO

			ENFERMO		Total
			Enfermos	No enfermos	
EXPOSICI	Expuestos	Recuento	22	1323	1345
		% de EXPOSICI	1,6%	98,4%	100,0%
	No expuestos	Recuento	12	1545	1557
		% de EXPOSICI	,8%	99,2%	100,0%
Total		Recuento	34	2868	2902
		% de EXPOSICI	1,2%	98,8%	100,0%

Estimación de riesgo

	Valor	Intervalo de confianza al 95%	
		Inferior	Superior
Razón de las ventajas para EXPOSICI (Expuestos / No expuestos)	2,141	1,056	4,343
Para la cohorte ENFERMO = Enfermos	2,122	1,054	4,272
Para la cohorte ENFERMO = No enfermos	,991	,983	,999
N de casos válidos	2902		

Definición de *odds ratio*

- El odds ratio (OR) se define como la razón del riesgo relativo de padecer la enfermedad respecto al riesgo relativo de no padecerla. Es decir:

$$OR = \frac{P(E / R) / P(E / \bar{R})}{P(\bar{E} / R) / P(\bar{E} / \bar{R})}$$

El odds ratio como aproximación del riesgo relativo

- Si la patología es poco prevalente se cumplirá:

$$P(E / R) \ll P(\bar{E} / R)$$

$$P(E / \bar{R}) \ll P(\bar{E} / \bar{R})$$

- Por lo tanto, en este caso:

$$OR = \frac{P(E / R) / P(E / \bar{R})}{P(\bar{E} / R) / P(\bar{E} / \bar{R})} \approx P(E / R) / P(E / \bar{R})$$

Ejemplo: El odds ratio como aproximación del riesgo relativo

Tabla de contingencia EXPOSICI * ENFERMO

			ENFERMO		Total
			Enfermos	No enfermos	
EXPOSICI	Expuestos	Recuento	100	100000	100100
		% de EXPOSICI	,100%	99,9%	100,0%
	No expuestos	Recuento	20	100000	100020
		% de EXPOSICI	,020%	100,0%	100,0%
Total		Recuento	120	200000	200120
		% de EXPOSICI	,060%	99,9%	100,0%

Estimación de riesgo

	Valor	Intervalo de confianza al 95%	
		Inferior	Superior
Razón de las ventajas para EXPOSICI (Expuestos / No expuestos)	5,000	3,093	8,082
Para la cohorte ENFERMO = Enfermos	4,996	3,091	8,074
Para la cohorte ENFERMO = No enfermos	,999	,999	,999
N de casos válidos	200120		

Cálculo del odds ratio

$$OR = \frac{P(E/R) / P(E/\bar{R})}{P(\bar{E}/R) / P(\bar{E}/\bar{R})} = \frac{P(E/R) \times P(\bar{E}/\bar{R})}{P(\bar{E}/R) \times P(E/\bar{R})}$$

$$OR = \frac{P(E \cap R) \times P(\bar{E} \cap \bar{R})}{P(\bar{E} \cap R) \times P(E \cap \bar{R})} \rightarrow \frac{f_{E \cap R} \times f_{\bar{E} \cap \bar{R}}}{f_{\bar{E} \cap R} \times f_{E \cap \bar{R}}} = \frac{15 \times 165}{5 \times 145} = 3.41$$

Tabla de contingencia EXPOSICI * ENFERMO

			ENFERMO		Total
			Enfermos	No enfermos	
EXPOSICI	Expuestos	Recuento	15	145	160
		% de EXPOSICI	9,4%	90,6%	100,0%
	No expuestos	Recuento	5	165	170
		% de EXPOSICI	2,9%	97,1%	100,0%
Total		Recuento	20	310	330
		% de EXPOSICI	6,1%	93,9%	100,0%

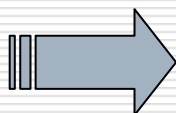
Es aconsejable utilizar el riesgo relativo en aquellas situaciones donde puede estimarse

Tabla de contingencia EXPOSICI * ENFERMO

			ENFERMO		Total
			Enfermos	No enfermos	
EXPOSICI	Expuestos	Recuento	23	145	168
		% de EXPOSICI	13,7%	86,3%	100,0%
	No expuestos	Recuento	12	165	177
		% de EXPOSICI	6,8%	93,2%	100,0%
Total		Recuento	35	310	345
		% de EXPOSICI	10,1%	89,9%	100,0%

Estimación de riesgo

	Valor	Intervalo de confianza al 95%	
		Inferior	Superior
Razón de las ventajas para EXPOSICI (Expuestos / No expuestos)	2,181	1,048	4,538
Para la cohorte ENFERMO = Enfermos	2,019	1,038	3,928
Para la cohorte ENFERMO = No enfermos	,926	,861	,995
N de casos válidos	345		



Ejemplo

- Se estudian los efectos de la contracepción oral (CO) en enfermedad cardiovascular en mujeres entre 40 y 44 años. En un conjunto de 5000 mujeres usuarias de CO, 13 presentaron infarto de miocardio en el periodo estudiado. En 10000 mujeres que no usaron CO en el mismo periodo, 7 presentaron infarto. Estima el riesgo relativo de padecer infarto de miocardio a partir de estos datos.

Ejemplo

Tabla de contingencia CO * IM

			IM		Total
			Si	No	
CO	Si	Recuento	13	4987	5000
		% de CO	,260%	99,7%	100,0%
	No	Recuento	7	9993	10000
		% de CO	,070%	99,9%	100,0%
Total		Recuento	20	14980	15000
		% de CO	,133%	99,9%	100,0%

Estimación de riesgo

	Valor	Intervalo de confianza al 95%	
		Inferior	Superior
Razón de las ventajas para CO (Si / No)	3,721	1,484	9,333
Para la cohorte IM = Si	3,714	1,483	9,304
Para la cohorte IM = No	,998	,997	1,000
N de casos válidos	15000		

Necesidad del planteamiento multivariante

Variable	OR (IC 95%)	
Edad (Incremento de 5 años)	1.07	2.23
Colesterol (Alto vs Bajo)	1.32	3.65
Sexo (Hombre vs Mujer)	0.53	0.78
Hipertensión	2.34	5.64
Hiperuricemia	1.90	8.34
Genotipo (AA vs. Aa)	0.23	0.57

- ❑ La edad, el colesterol, el género, la hipertensión y la hiperuricemia son factores de riesgo
- ❑ ¿Cómo podemos estimar el riesgo relativo en función de estos factores?
- ❑ ¿Qué factores son más significativos?

Necesidad del planteamiento multivariante

- ❑ Considerar simultáneamente el efecto de distintas variables
- ❑ Seleccionar las variables más significativas
- ❑ Estimar riesgos relativos ajustados según determinados valores de las variables consideradas

❑ *Regresión logística*

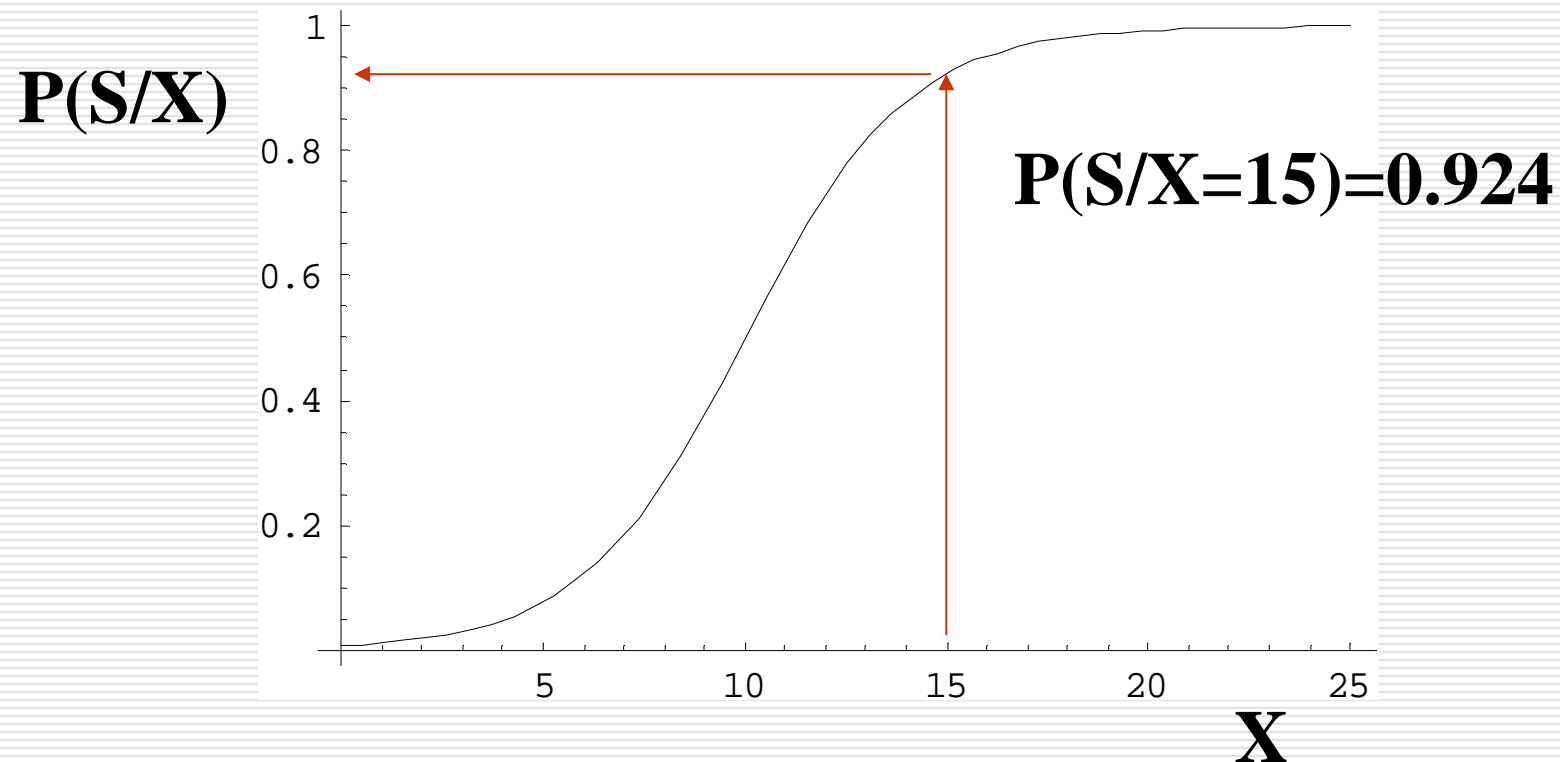
Introducción a la regresión logística

El modelo logístico para la probabilidad de un suceso

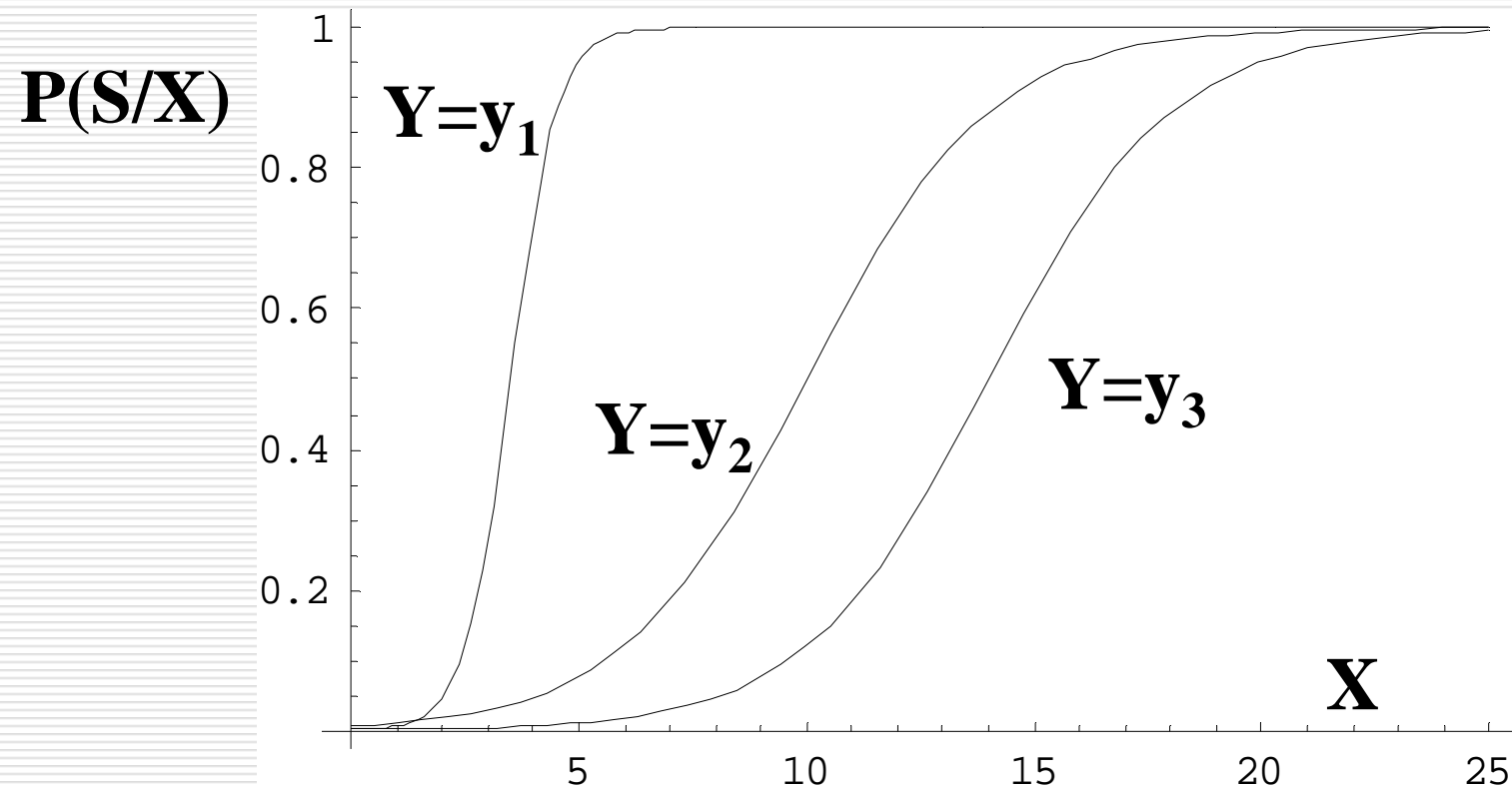
- ¿Cómo podemos indicar la probabilidad de que se presente un determinado suceso en función de los valores de distintas variables?
 - La probabilidad debe estar entre 0 y 1
 - Distintos valores de las variables deben proporcionar distintas probabilidades
 - La probabilidad es creciente (o de creciente) en función de los valores de la variable

El modelo logístico

Probabilidad de un suceso en función de los valores de una variable

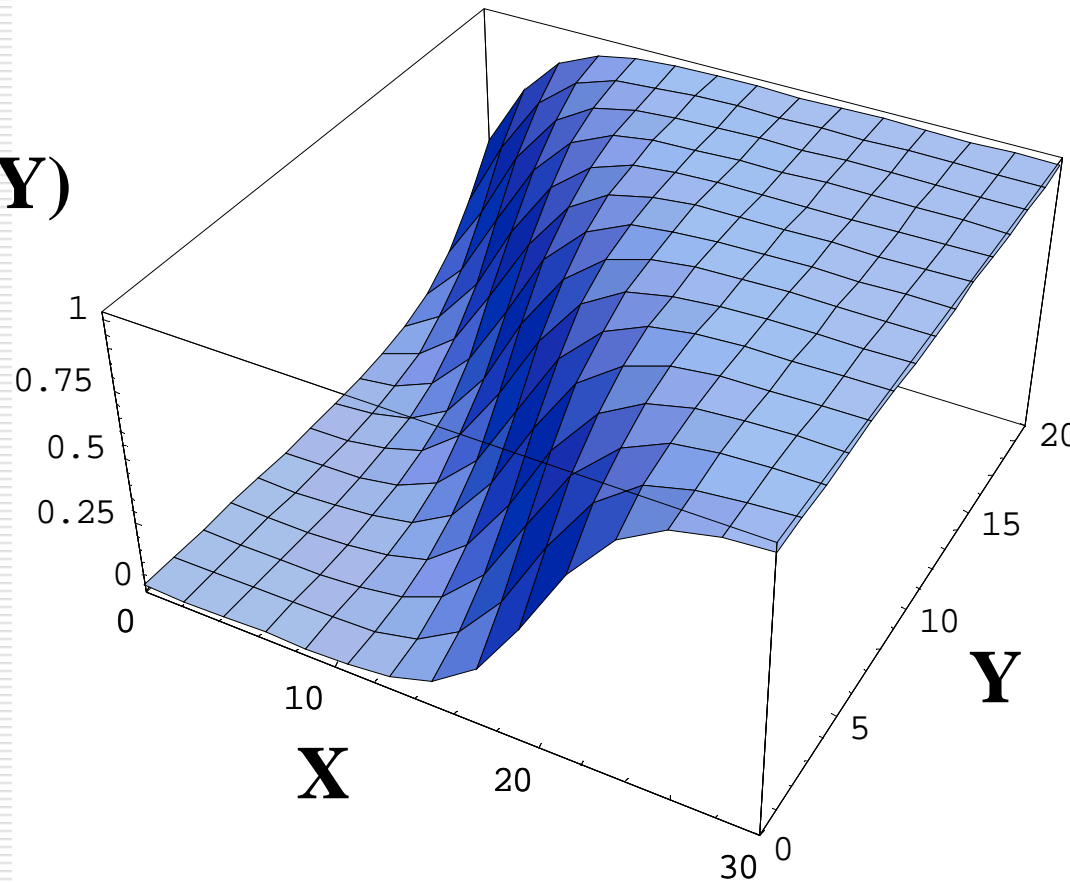


$P(S/X)$ puede cambiar en función del grupo o de terceras variables



$P(S/X)$ puede cambiar en función del grupo o de terceras variables

$P(S/X, Y)$



El modelo logístico univariante

$$P(S / X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}}$$

$$\frac{P(S / X)}{P(\bar{S} / X)} = e^{(\beta_0 + \beta_1 X)}$$

$$\text{Log} \left(\frac{P(S / X)}{P(\bar{S} / X)} \right) = \beta_0 + \beta_1 X$$

El modelo logístico univariante

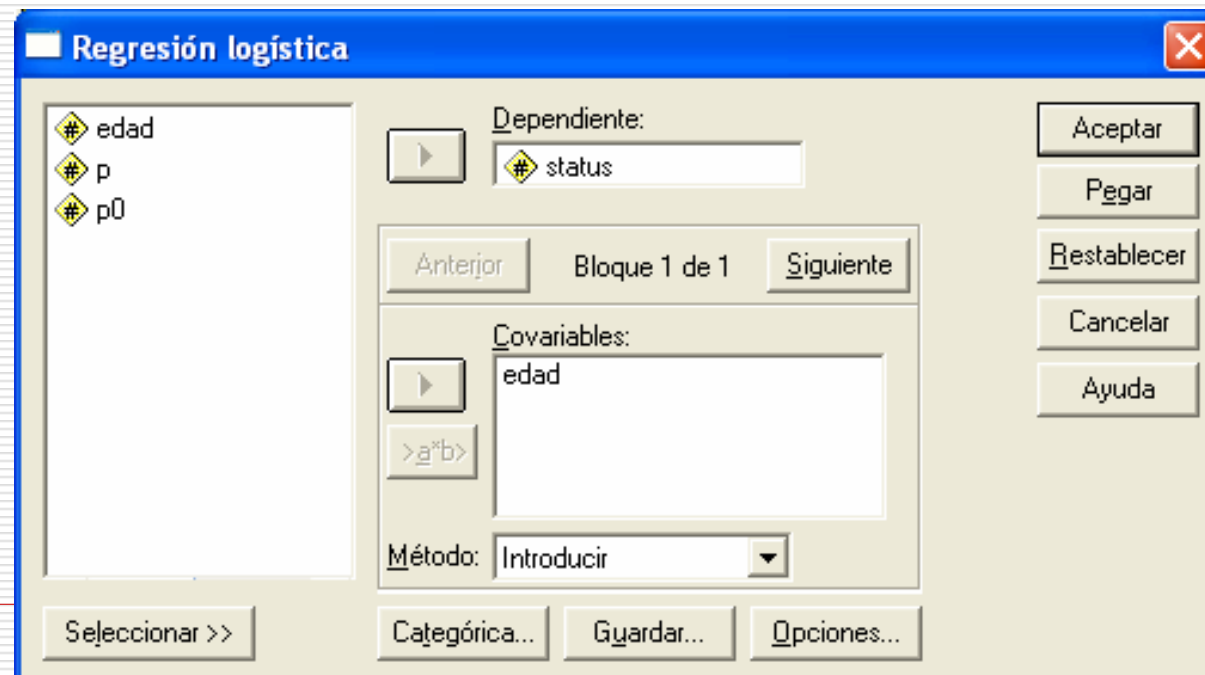
$$\frac{P(S / X = x_2)}{P(\bar{S} / X = x_2)} \cdot \frac{P(S / X = x_1)}{P(\bar{S} / X = x_1)} = \frac{P(S / X = x_2)P(\bar{S} / X = x_1)}{P(\bar{S} / X = x_2)P(S / X = x_1)} =$$
$$\frac{P(S / X = x_2) / P(\bar{S} / X = x_2)}{P(S / X = x_1) / P(\bar{S} / X = x_1)} = OR$$

$$\left. \begin{aligned} \frac{P(S / X = x_1)}{P(\bar{S} / X = x_1)} &= e^{(\beta_0 + \beta_1 x_1)} \\ \frac{P(S / X = x_2)}{P(\bar{S} / X = x_2)} &= e^{(\beta_0 + \beta_1 x_2)} \end{aligned} \right\} \Rightarrow OR = e^{\beta_1 (x_2 - x_1)}$$

Odds ratio (estimación del riesgo relativo) asociado a un cambio de x_2 a x_1

Ejemplo

- Disponemos de la edad de cada individuo y determinamos si ha padecido o no una determinada complicación relacionada con un tratamiento (Variable Status: 1(Si) 0 (No))



Ejemplo

Codificación de la variable dependiente

Valor original	Valor interno
,00	0
1,00	1

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95,0% para EXP(B)	
								Inferior	Superior
Paso a 1	EDAD	,244	,059	17,289	1	,000	1,276	1,138	1,432
	Constante	-10,579	2,796	14,318	1	,000	,000		

a. Variable(s) introducida(s) en el paso 1: EDAD.

$$P(S / edad) = \frac{1}{1 + e^{-(-10.58 + 0.24 * edad)}} \quad OR = e^{0.24 * (e_2 - e_1)}$$

$$OR = e^{0.24} = 1.276$$

Una persona que tiene un año más tiene un riesgo relativo de 1.3 respecto a una persona que tiene un año menos

$$OR = e^{0.24 * 5} = 3.32$$

Una persona que tiene cinco años más tiene un riesgo relativo de 3.3 respecto a una persona que tiene cinco años menos

Tratamiento de variables discretas (nominales)

Tabla de contingencia HIPERT * STATUS

			STATUS		Total
			Si	No	
HIPERT	Si	Recuento	28	10	38
		% de HIPERT	73,7%	26,3%	100,0%
	No	Recuento	23	10	33
		% de HIPERT	69,7%	30,3%	100,0%
Total		Recuento	51	20	71
		% de HIPERT	71,8%	28,2%	100,0%

Estimación de riesgo

	Valor	Intervalo de confianza al 95%	
		Inferior	Superior
Razón de las ventajas para HIPERT (Si / No)	1,217	,432	3,429
Para la cohorte STATUS = Si	1,057	,788	1,419
Para la cohorte STATUS = No	,868	,413	1,824
N de casos válidos	71		

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95,0% para EXP(B)	
								Inferior	Superior
Paso 1 ^a	HIPERT(1)	,197	,528	,139	1	,710	1,217	,432	3,429
	Constante	,833	,379	4,835	1	,028	2,300		

a. Variable(s) introducida(s) en el paso 1: HIPERT.

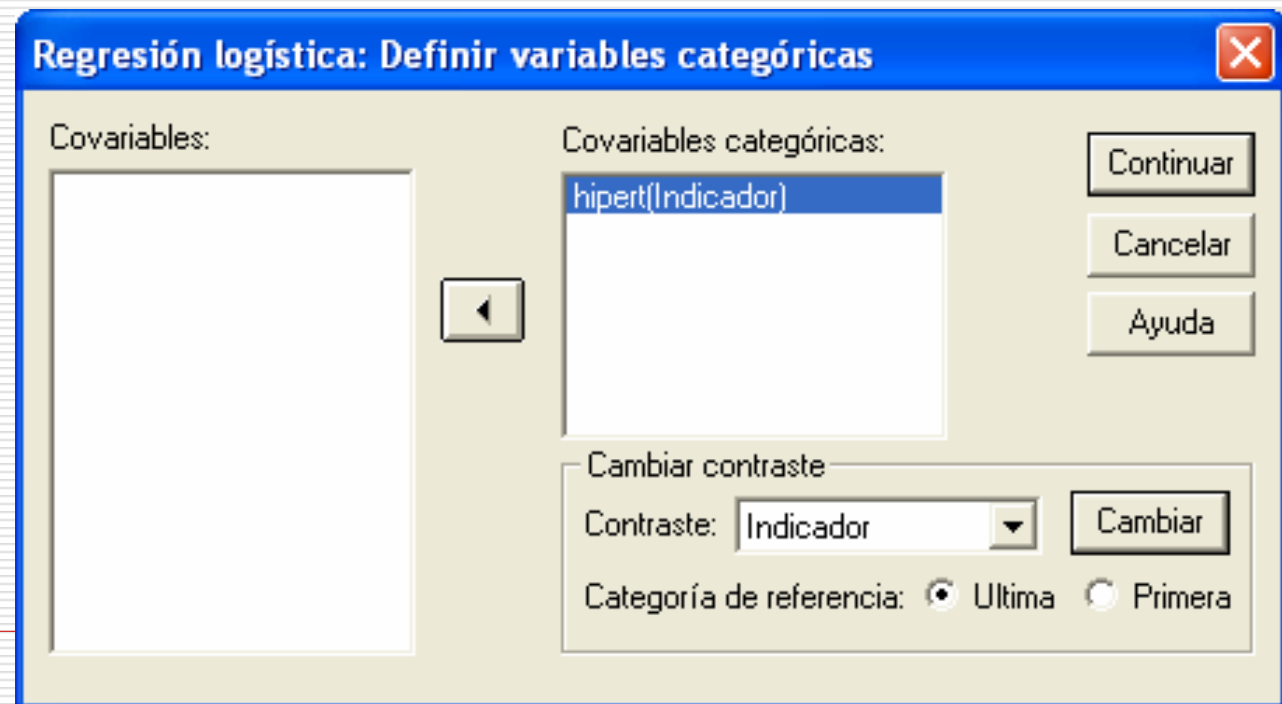
Tratamiento de variables discretas (nominales)

Hipertensión: 1 (Si) 2 (No)

Categoría de referencia es NO

Especificar que el último valor es el de referencia !!

Status: 1 (Si) 0 (No)



Tratamiento de variables discretas (nominales)

Tabla de contingencia GENO * STATUS

			STATUS		Total
			No	Si	
GENO	aa	Recuento	23	13	36
		% de GENO	63,9%	36,1%	100,0%
	Aa	Recuento	16	22	38
		% de GENO	42,1%	57,9%	100,0%
	AA	Recuento	7	33	40
		% de GENO	17,5%	82,5%	100,0%
Total	Recuento	46	68	114	
	% de GENO	40,4%	59,6%	100,0%	

Codificaciones de variables categóricas

		Frecuencia	Codificación de	
			(1)	(2)
GENO	aa	2	1,000	,000
	Aa	2	,000	1,000
	AA	2	,000	,000

aa : 1

Aa : 2

AA : 3

Categoría de Referencia: Última

$$\frac{13 \times 7}{23 \times 33} = 0.12$$

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1 a	GENO			15,328	2	,000	
	GENO(1)	-2,121	,542	15,326	1	,000	,120
	GENO(2)	-1,232	,530	5,400	1	,020	,292
	Constante	1,550	,416	13,884	1	,000	4,714

a. Variable(s) introducida(s) en el paso 1: GENO.

Tratamiento de variables discretas (nominales)

Tabla de contingencia GENO * STATUS

			STATUS		Total
			No	Si	
GENO	aa	Recuento	23	13	36
		% de GENO	63,9%	36,1%	100,0%
	Aa	Recuento	16	22	38
		% de GENO	42,1%	57,9%	100,0%
	AA	Recuento	7	33	40
		% de GENO	17,5%	82,5%	100,0%
Total	Recuento	46	68	114	
	% de GENO	40,4%	59,6%	100,0%	

Codificaciones de variables categóricas

		Frecuencia	Codificación de	
			(1)	(2)
GENO	aa	2	,000	,000
	Aa	2	1,000	,000
	AA	2	,000	1,000

aa : 1
Aa : 2
AA : 3

Categoría de Referencia: Primera

$$\frac{33 \times 23}{7 \times 13} = 8.34$$

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1 ^a	GENO			15,328	2	,000	
	GENO(1)	,889	,478	3,461	1	,063	2,433
	GENO(2)	2,121	,542	15,326	1	,000	8,340
	Constante	-,571	,347	2,704	1	,100	,565

a. Variable(s) introducida(s) en el paso 1: GENO.

El modelo logístico multivariante

$$X = \{ X_1, X_2, \dots, X_n \}$$

$$P(S / X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}}$$

$$\frac{P(S / X)}{P(\bar{S} / X)} = e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}$$

$$\text{Log} \left(\frac{P(S / X)}{P(\bar{S} / X)} \right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

$$\frac{P(S / X_i = x_i)}{P(\bar{S} / X_i = x_i)} = e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_i x_i + \beta_n X_n)}$$

$$\frac{P(S / X_i = x_j)}{P(\bar{S} / X_i = x_j)} = e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_i x_j + \beta_n X_n)}$$

$$OR = \frac{P(S / X_i = x_i) / P(\bar{S} / X_i = x_i)}{P(S / X_i = x_j) / P(\bar{S} / X_i = x_j)} = e^{\beta_i (x_i - x_j)}$$

OR correspondiente a la estimación del RR cuando X pasa de x_i a x_j

Ejemplo

- Disponemos de datos acerca de hombres y mujeres en relación a la aparición de molestias artríticas. Queremos establecer la posible relación de estas molestias con la concentración dos metabolitos (X_1, X_2).

Resultados

Codificaciones de variables categóricas

		Frecuencia	Codificación de
			(1)
SEXO	Hombre	59	1,000
	Mujer	52	,000

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95,0% para EXP(B)	
								Inferior	Superior
Paso a 1	SEXO(1)	,451	,418	1,164	1	,281	1,569	,692	3,559
	X1	,253	,110	5,277	1	,022	1,288	1,038	1,597
	X2	,319	,114	7,852	1	,005	1,376	1,101	1,721
	Constante	-8,086	2,697	8,988	1	,003	,000		

a. Variable(s) introducida(s) en el paso 1: SEXO, X1, X2.

Interpretación

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95,0% para EXP(B)	
								Inferior	Superior
Paso a 1	SEXO(1)	,451	,418	1,164	1	,281	1,569	,692	3,559
	X1	,253	,110	5,277	1	,022	1,288	1,038	1,597
	X2	,319	,114	7,852	1	,005	1,376	1,101	1,721
	Constante	-8,086	2,697	8,988	1	,003	,000		

a. Variable(s) introducida(s) en el paso 1: SEXO, X1, X2.

- La estimación del riesgo relativo de hombres respecto a mujeres es:

$$OR(H / M) = 1.57$$

- Esta estimación está ajustada por las variables X1 y X2

Interpretación

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95,0% para EXP(B)	
								Inferior	Superior
Paso a 1	SEXO(1)	,451	,418	1,164	1	,281	1,569	,692	3,559
	X1	,253	,110	5,277	1	,022	1,288	1,038	1,597
	X2	,319	,114	7,852	1	,005	1,376	1,101	1,721
	Constante	-8,086	2,697	8,988	1	,003	,000		

a. Variable(s) introducida(s) en el paso 1: SEXO, X1, X2.



Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95,0% para EXP(B)	
								Inferior	Superior
Paso a 1	X2	,293	,109	7,205	1	,007	1,340	1,082	1,660
	Constante	-1,823	,630	8,383	1	,004	,162		

a. Variable(s) introducida(s) en el paso 1: X2.

Selección de variables

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95,0% para EXP(B)	
								Inferior	Superior
Paso a 1	SEXO(1)	,451	,418	1,164	1	,281	1,569	,692	3,559
	X1	,253	,110	5,277	1	,022	1,288	1,038	1,597
	X2	,319	,114	7,852	1	,005	1,376	1,101	1,721
	Constante	-8,086	2,697	8,988	1	,003	,000		

a. Variable(s) introducida(s) en el paso 1: SEXO, X1, X2.

❑ Podemos eliminar las variables que no son significativas

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95,0% para EXP(B)	
								Inferior	Superior
Paso a 1	X1	,267	,110	5,917	1	,015	1,306	1,053	1,619
	X2	,309	,113	7,501	1	,006	1,362	1,092	1,699
	Constante	-8,111	2,701	9,019	1	,003	,000		

a. Variable(s) introducida(s) en el paso 1: X1, X2.

Modelo final

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95,0% para EXP(B)	
								Inferior	Superior
Paso a 1	X1	,267	,110	5,917	1	,015	1,306	1,053	1,619
	X2	,309	,113	7,501	1	,006	1,362	1,092	1,699
	Constante	-8,111	2,701	9,019	1	,003	,000		

a. Variable(s) introducida(s) en el paso 1: X1, X2.

$$P(S / X) = \frac{1}{1 + e^{-(-8.11 + 0.267 X_1 + 0.309 X_2)}}$$

$$OR\left(\frac{X_1 \rightarrow \Delta x_1}{X_2 \rightarrow \Delta x_2}\right) = e^{(\beta_1 \Delta x_1 + \beta_2 \Delta x_2)} \Rightarrow e^{(0.267(23-21) + 0.309(6-5.5))} = 1.99$$

Uso de la regresión logística como método de análisis de la relación entre variables cualitativas

- ❑ A menudo queremos evaluar si existe relación entre distintas variables cualitativas (tablas múltiples)
- ❑ Cuando el objetivo es calcular la probabilidad de un suceso en función de varias variables, podemos utilizar la regresión logística
- ❑ P.e. Probabilidad de complicaciones en función de la gravedad (+, ++, +++), el sexo (Hombre/Mujer), y el genotipo (aa, Aa, AA).

Uso de la regresión logística como método de análisis de la relación entre variables cualitativas

Codificaciones de variables categóricas

	Frecuencia	Codificación de	
		(1)	(2)
GRAVEDAD +	14	,000	,000
++	22	1,000	,000
+++	6	,000	1,000
GENOTIPO AA	14	,000	,000
Aa	11	1,000	,000
aa	17	,000	1,000
SEXO Hombre	23	1,000	
Mujer	19	,000	

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95,0% para EXP(B)	
								Inferior	Superior
Paso a 1	GENOTIPO			1,258	2	,533			
	GENOTIPO(1)	,691	,955	,524	1	,469	1,996	,307	12,964
	GENOTIPO(2)	-,330	,797	,171	1	,679	,719	,151	3,433
	SEXO(1)	1,123	,718	2,443	1	,118	3,074	,752	12,563
	GRAVEDAD			2,025	2	,363			
	GRAVEDAD(1)	1,076	,756	2,025	1	,155	2,934	,666	12,922
	GRAVEDAD(2)	,639	1,130	,320	1	,572	1,895	,207	17,364
	Constante	-,874	,939	,865	1	,352	,417		

a. Variable(s) introducida(s) en el paso 1: GENOTIPO, SEXO, GRAVEDAD.

Uso de la regresión logística como método de análisis de la relación entre variables cualitativas

Codificaciones de variables categóricas

	Frecuencia	Codificación de	
		(1)	(2)
GRAVEDAD +	14	,000	,000
++	22	1,000	,000
+++	6	,000	1,000
GENOTIPO AA	14	,000	,000
Aa	11	1,000	,000
aa	17	,000	1,000
SEXO Hombre	23	1,000	
Mujer	19	,000	

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95,0% para EXP(B)	
								Inferior	Superior
Paso a 1	GENOTIPO			1,753	2	,416			
	GENOTIPO(1)	,688	,603	1,303	1	,254	1,991	,610	6,491
	GENOTIPO(2)	,722	,602	1,435	1	,231	2,058	,632	6,702
	SEXO(1)	-,176	,473	,139	1	,709	,838	,332	2,117
	GRAVEDAD			13,566	2	,001			
	GRAVEDAD(1)	1,123	,514	4,766	1	,029	3,074	1,122	8,425
	GRAVEDAD(2)	2,547	,711	12,841	1	,000	12,775	3,171	51,463
	Constante	-1,284	,615	4,360	1	,037	,277		

a. Variable(s) introducida(s) en el paso 1: GENOTIPO, SEXO, GRAVEDAD.

La regresión logística como alternativa al procedimiento de Mantel-Haenzel

Tabla de contingencia LOW * SMOKE * RACE

Recuento			SMOKE		Total
RACE			,00	1,00	
1,00	LOW	,00	40	33	73
		1,00	4	19	23
	Total		44	52	96
2,00	LOW	,00	11	4	15
		1,00	5	6	11
	Total		16	10	26
3,00	LOW	,00	35	7	42
		1,00	20	5	25
	Total		55	12	67

Estimación de riesgo

RACE		Valor	Intervalo de confianza al 95%	
			Inferior	Superior
1,00	Razón de las ventajas para LOW (,00 / 1,00)	5,758	1,782	18,599
	N de casos válidos	96		
2,00	Razón de las ventajas para LOW (,00 / 1,00)	3,300	,635	17,160
	N de casos válidos	26		
3,00	Razón de las ventajas para LOW (,00 / 1,00)	1,250	,350	4,462
	N de casos válidos	67		

La regresión logística como alternativa al procedimiento de Mantel-Haenzel

Odds ratios ajustados

Estimación de riesgo

RACE		Valor	Intervalo de confianza al 95%	
			Inferior	Superior
1,00	Razón de las ventajas para LOW (.00 / 1,00) N de casos válidos	5,758 96	1,782	18,599
2,00	Razón de las ventajas para LOW (.00 / 1,00) N de casos válidos	3,300 26	,635	17,160
3,00	Razón de las ventajas para LOW (.00 / 1,00) N de casos válidos	1,250 67	,350	4,462

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95,0% para EXP(B)	
								Inferior	Superior
Paso 1	SMOKE(1)	1,116	,369	9,135	1	,003	3,052	1,480	6,294
	RACE			9,112	2	,011			
	RACE(1)	1,084	,490	4,894	1	,027	2,956	1,132	7,724
	RACE(2)	1,108	,400	7,668	1	,006	3,030	1,382	6,639
	Constante	-1,840	,353	27,205	1	,000	,159		

a. Variable(s) introducida(s) en el paso 1: SMOKE, RACE.

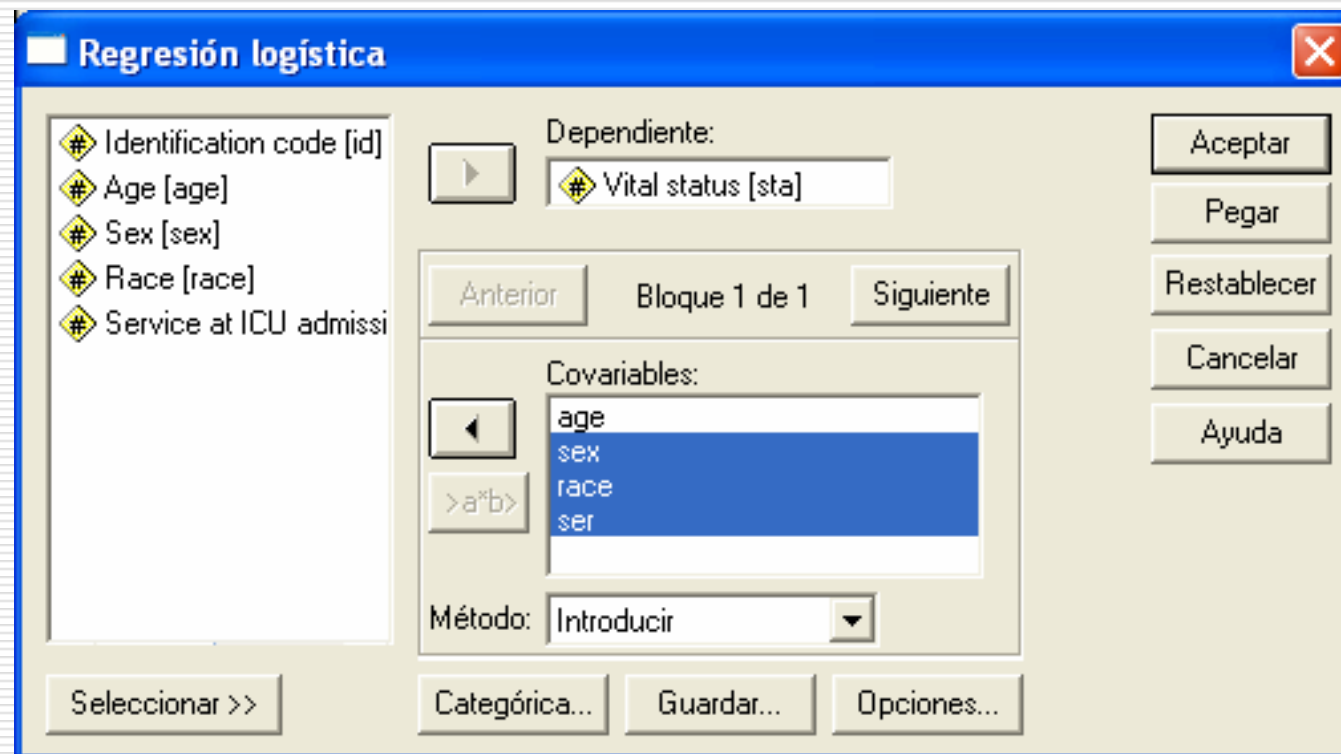
Modelos con variables cualitativas y cuantitativas

- ❑ En un mismo modelo de regresión logística podemos mezclar variables cualitativas (nominales u ordinales) y variables cuantitativas
- ❑ La codificación de variables cualitativas debe hacerse con cuidado para facilitar la interpretación de resultados
- ❑ En las variables cualitativas debemos escoger una categoría de referencia para el cálculo de *odds ratios*.

Datos de evolución en ICU

- ☐ Edad
- ☐ Sexo (0: Male, 1: Female)
- ☐ Race (1: White, 2: Black, 3: Other)
- ☐ Service (0: Medical, Surgical)

SPSS



SPSS

Regresión logística: Definir variables categóricas

Covariables:
Age [age]

Covariables categóricas:
sex(Indicador)
race(Indicador)
ser(Indicador)

Cambiar contraste
Contraste: Indicador
Categoría de referencia: ☒ Última ☐ Primera

Continuar
Cancelar
Ayuda

Codificaciones de variables categóricas

		Frecuencia	Codificación de	
			(1)	(2)
Race	White	175	1,000	,000
	Black	15	,000	1,000
	Other	10	,000	,000
Service at ICU admission	Medical	93	1,000	
	Surgical	107	,000	
Sex	Male	124	1,000	
	Female	76	,000	

Race(1): White
Race(2): Black

Ser(1): Medical

Sex(1): Male

SPSS

Race(1): White
Race(2): Black

Ser(1): Medical

Sex(1): Male

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95,0% para EXP(B)	
								Inferior	Superior
Paso 1 ^a	AGE	,028	,011	6,256	1	,012	1,029	1,006	1,052
	SEX(1)	,028	,379	,005	1	,941	1,028	,490	2,160
	RACE			1,493	2	,474			
	RACE(1)	-,395	,852	,215	1	,643	,673	,127	3,580
	RACE(2)	-1,579	1,331	1,407	1	,236	,206	,015	2,801
	SER(1)	1,071	,382	7,885	1	,005	2,919	1,382	6,167
	Constante	-3,272	1,059	9,551	1	,002	,038		

a. Variable(s) introducida(s) en el paso 1: AGE, SEX, RACE, SER.

- La admisión en un servicio médico determina una probabilidad más elevada de muerte (OR: 1.38 – 6.17)
- La edad se asocia significativamente con una mayor probabilidad de muerte
- El sexo y el grupo étnico no se relacionan significativamente con la probabilidad de muerte

ICU.sav - Editor de datos SPSS

Archivo Edición Ver Datos Transformar Analizar Gráficos Utilidades Ventana ?

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores
1	id	Numérico	3	0	Identification code	Ninguno
2	sta	Numérico	1	0	Vital status	{0, Lived}...
3	age	Numérico	2	0	Age	Ninguno
4	sex	Numérico	1	0	Sex	{0, Male}...
5	race	Numérico	1	0	Race	{1, White}...
6	ser	Numérico	1	0	Service at ICU admission	{0, Medical}...
7	can	Numérico	1	0	Cancer part of present problem	{0, No}...
8	crn	Numérico	1	0	History of chronic renal failure	{0, No}...
9	inf	Numérico	1	0	Infection probable at ICU admission	{0, No}...
10	cpr	Numérico	1	0	CPR prior to ICU admission	{0, No}...
11	sys	Numérico	3	2	Systolic blood pressure at ICU admission	Ninguno
12	hra	Numérico	3	2	Heart rate at ICU admission	Ninguno
13	pre	Numérico	1	0	Previous admission to an ICU within 6 months	{0, No}...
14	typ	Numérico	1	0	Type of admission	{0, Elective}...
15	fra	Numérico	1	0	Long bone, Multiple, Neck, Single area, or Hip fracture	{0, No}...
16	po2	Numérico	1	0	PO2 from initial blood gases	{0, >60}...
17	ph	Numérico	1	0	PH from initial blood gases	{0, >=7.25}...
18	pco	Numérico	1	0	PCO2 from initial blood gases	{0, <=45}...
19	bic	Numérico	1	0	Bicarbonate from initial blood gases	{0, >=18}...
20	cre	Numérico	1	0	Creatinine from initial blood gases	{0, <=2.0}...
21	loc	Numérico	1	0	Level of conciosness at ICU admission	{0, No coma or stupor }...

Vista de datos Vista de variables

SPSS El procesador está preparado

Codificación de variables

Codificaciones de variables categóricas

		Frecuencia	Codificación de	
			(1)	(2)
Level of conciosness at ICU admission	No coma or stupor	185	1,000	,000
	Depp stupor	5	,000	1,000
	Coma	10	,000	,000
Race	White	175	1,000	,000
	Black	15	,000	1,000
	Other	10	,000	,000
Service at ICU admission	Medical	93	1,000	
	Surgical	107	,000	
Cancer part of present problem	No	180	1,000	
	Yes	20	,000	
History of chronic renal failure	No	181	1,000	
	Yes	19	,000	
Infection probable at ICU admission	No	116	1,000	
	Yes	84	,000	
CPR prior to ICU admission	No	187	1,000	
	Yes	13	,000	
Previous admission to an ICU within 6 months	No	170	1,000	
	Yes	30	,000	
Type of admission	Elective	53	1,000	
	Emergency	147	,000	
Creatinine from initial blood gases	<=2.0	190	1,000	
	>2.0	10	,000	
Bicarbonate from initial blood gases	>=18	185	1,000	
	<18	15	,000	
PCO2 from initial blood gases	<=45	180	1,000	
	>45	20	,000	
PH from initial boood gases	>=7.25	187	1,000	
	<7.25	13	,000	
Long bone, Multiple, Neck, Single area, or rib fracture	No	185	1,000	
	Yes	15	,000	
PO2 from initial blood gases	>60	184	1,000	
	<=60	16	,000	
Sex	Male	124	1,000	
	Female	76	,000	

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95,0% para EXP(B)	
								Inferior	Superior
Paso 1 ^a	AGE	,056	,018	9,331	1	,002	1,058	1,020	1,097
	SEX(1)	,721	,546	1,746	1	,186	2,057	,706	5,999
	RACE			,310	2	,856			
	RACE(1)	-,583	1,313	,197	1	,657	,558	,043	7,313
	RACE(2)	-7,438	20,543	,131	1	,717	,001	,000	1,8E+14
	SER(1)	,674	,629	1,148	1	,284	1,962	,572	6,729
	CAN(1)	-3,483	1,121	9,650	1	,002	,031	,003	,277
	CRN(1)	-,119	,845	,020	1	,888	,888	,170	4,649
	INF(1)	,108	,556	,038	1	,846	1,114	,375	3,311
	CPR(1)	-1,032	,990	1,087	1	,297	,356	,051	2,480
	SYS	-,021	,009	4,871	1	,027	,979	,961	,998
	HRA	-,003	,010	,080	1	,778	,997	,977	1,017
	PRE(1)	-1,279	,702	3,321	1	,068	,278	,070	1,101
	TYP(1)	-3,748	1,342	7,798	1	,005	,024	,002	,327
	FRA(1)	-1,649	1,093	2,277	1	,131	,192	,023	1,637
	PO2(1)	,677	,940	,518	1	,472	1,967	,312	12,419
	PH(1)	-1,771	1,212	2,134	1	,144	,170	,016	1,832
	PCO(1)	2,084	1,165	3,201	1	,074	8,033	,820	78,741
	BIC(1)	,262	,897	,086	1	,770	1,300	,224	7,537
	CRE(1)	-,100	1,131	,008	1	,929	,904	,099	8,296
	LOC			6,857	2	,032			
	LOC(1)	-3,458	1,341	6,646	1	,010	,031	,002	,436
	LOC(2)	15,659	39,484	,157	1	,692	6321447	,000	2,57E+40
	Constante	7,149	3,139	5,188	1	,023	1273,250		

a. Variable(s) introducida(s) en el paso 1: AGE, SEX, RACE, SER, CAN, CRN, INF, CPR, SYS, HRA, PRE, TYP, FRA, PO2, PH, PCO, BIC, CRE, LOC.

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95,0% para EXP(B)	
								Inferior	Superior
Paso 1 ^a	AGE	,042	,014	8,402	1	,004	1,043	1,014	1,073
	SEX(1)	,484	,456	1,130	1	,288	1,623	,665	3,964
	RACE			,729	2	,694			
	RACE(1)	-,221	,987	,050	1	,823	,802	,116	5,550
	RACE(2)	-1,130	1,433	,622	1	,430	,323	,019	5,353
	SER(1)	,349	,533	,429	1	,512	1,418	,499	4,030
	CAN(1)	-2,083	,908	5,267	1	,022	,125	,021	,738
	CRN(1)	-,314	,676	,216	1	,642	,730	,194	2,745
	INF(1)	-,172	,463	,139	1	,709	,842	,340	2,085
	CPR(1)	-1,548	,752	4,237	1	,040	,213	,049	,929
	SYS	-,013	,007	3,629	1	,057	,987	,973	1,000
	HRA	-,011	,009	1,491	1	,222	,989	,972	1,007
	PRE(1)	-,728	,588	1,536	1	,215	,483	,153	1,527
	TYP(1)	-2,724	,940	8,403	1	,004	,066	,010	,414
	FRA(1)	-,651	,949	,471	1	,493	,521	,081	3,349
	PO2(1)	-,207	,805	,066	1	,797	,813	,168	3,936
	PH(1)	-,557	,954	,341	1	,559	,573	,088	3,719
	PCO(1)	,949	,891	1,133	1	,287	2,582	,450	14,807
	BIC(1)	,271	,786	,119	1	,730	1,312	,281	6,120
	CRE(1)	-,446	,933	,229	1	,632	,640	,103	3,984
	Constante	3,953	2,554	2,396	1	,122	52,101		

a. Variable(s) introducida(s) en el paso 1: AGE, SEX, RACE, SER, CAN, CRN, INF, CPR, SYS, HRA, PRE, TYP, FRA, PO2, PH, PCO, BIC, CRE.

Selección de modelos

Regresión logística

Dependiente: Vital status [sta]

Anterior Bloque 1 de 1 Siguiente

Covariables:

age
sex(Cat)
race(Cat)
ser(Cat)
can(Cat)

Método: Adelante:Wald

Seleccionar >> Categórica... Guardar... Opciones...

Aceptar
Pegar
Restablecer
Cancelar
Ayuda

Selección de modelos

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95,0% para EXP(B)	
								Inferior	Superior
Paso 1	TYP(1)	-2,185	,745	8,600	1	,003	,112	,026	,484
	Constante	-1,054	,188	31,288	1	,000	,349		
Paso 2	AGE	,034	,011	10,117	1	,001	1,035	1,013	1,057
	TYP(1)	-2,454	,753	10,629	1	,001	,086	,020	,376
	Constante	-3,055	,693	19,453	1	,000	,047		
Paso 3	AGE	,035	,011	9,970	1	,002	1,036	1,013	1,059
	CPR(1)	-1,391	,616	5,093	1	,024	,249	,074	,833
	TYP(1)	-2,306	,757	9,286	1	,002	,100	,023	,439
	Constante	-1,885	,864	4,764	1	,029	,152		
Paso 4	AGE	,037	,012	10,561	1	,001	1,038	1,015	1,062
	CAN(1)	-1,534	,802	3,655	1	,056	,216	,045	1,039
	CPR(1)	-1,344	,616	4,766	1	,029	,261	,078	,872
	TYP(1)	-2,849	,867	10,800	1	,001	,058	,011	,317
	Constante	-,609	1,094	,310	1	,578	,544		
Paso 5	AGE	,037	,012	10,307	1	,001	1,038	1,015	1,062
	CAN(1)	-1,624	,808	4,033	1	,045	,197	,040	,962
	CPR(1)	-1,247	,629	3,933	1	,047	,287	,084	,986
	SYS	-,013	,006	4,434	1	,035	,987	,975	,999
	TYP(1)	-2,746	,879	9,756	1	,002	,064	,011	,359
	Constante	1,029	1,356	,576	1	,448	2,799		

Modelo final

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95,0% para EXP(B)	
								Inferior	Superior
Paso 1	AGE	,037	,012	10,307	1	,001	1,038	1,015	1,062
	CAN(1)	-1,624	,808	4,033	1	,045	,197	,040	,962
	CPR(1)	-1,247	,629	3,933	1	,047	,287	,084	,986
	SYS	-,013	,006	4,434	1	,035	,987	,975	,999
	TYP(1)	-2,746	,879	9,756	1	,002	,064	,011	,359
	Constante	1,029	1,356	,576	1	,448	2,799		

a. Variable(s) introducida(s) en el paso 1: AGE, CAN, CPR, SYS, TYP.

Codificaciones de variables categóricas

			Codificaci ón de
		Frecuencia	(1)
Type of admission	Elective	53	1,000
	Emergency	147	,000
CPR prior to ICU admission	No	187	1,000
	Yes	13	,000
Cancer part of present problem	No	180	1,000
	Yes	20	,000

Modelo final

Cambio de codificación

Cuidado: Los intervalos de confianza son muy amplios

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95,0% para EXP(B)	
								Inferior	Superior
Paso 1 ^a	AGE	,037	,012	10,307	1	,001	1,038	1,015	1,062
	CAN(1)	1,624	,808	4,033	1	,045	5,071	1,040	24,732
	CPR(1)	1,247	,629	3,933	1	,047	3,481	1,015	11,944
	SYS	-,013	,006	4,434	1	,035	,987	,975	,999
	TYP(1)	2,746	,879	9,756	1	,002	15,585	2,782	87,323
	Constante	-4,588	1,431	10,276	1	,001	,010		

a. Variable(s) introducida(s) en el paso 1: AGE, CAN, CPR, SYS, TYP.

Codificaciones de variables categóricas

			Codificaci ón de
		Frecuencia	(1)
Type of admission	Elective	53	,000
	Emergency	147	1,000
CPR prior to ICU admission	No	187	,000
	Yes	13	1,000
Cancer part of present problem	No	180	,000
	Yes	20	1,000

CAN, CPR y TYP son las variables más importantes

Interpretación

- ❑ Un individuo de 55 años ingresado en urgencias, con una presión sistólica de 100, con cancer y CPR previo tiene una probabilidad de morir igual a 0.85
- ❑ Un individuo de 55 años ingresado en urgencias, con una presión sistólica de 100, sin cancer ni CPR previo tiene una probabilidad de morir igual a 0.25
- ❑ Un individuo de 55 años ingresado en urgencias, con una presión sistólica de 100, con cancer y sin CPR previo tiene una probabilidad de morir igual a 0.62
- ❑ Un individuo de 55 años ingresado en urgencias, con una presión sistólica de 60, con cancer y sin CPR previo tiene una probabilidad de morir igual a 0.74

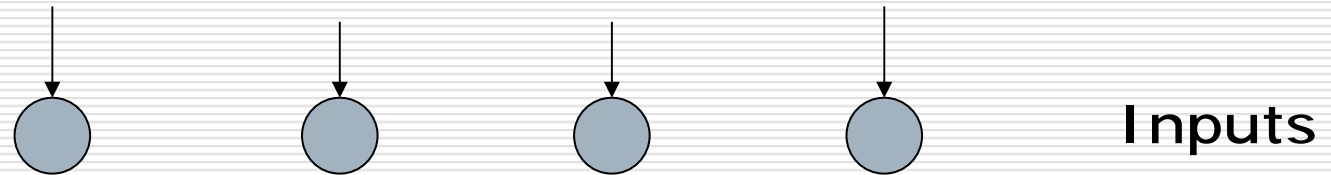
Limitaciones de la regresión logística

- Independencia de variables
 - Podemos considerar efectos de interacción entre variables
 - Es difícil concretar los efectos
- Efectos lineales de las variables en el valor del logit
- Alternativas: Redes Neuronales

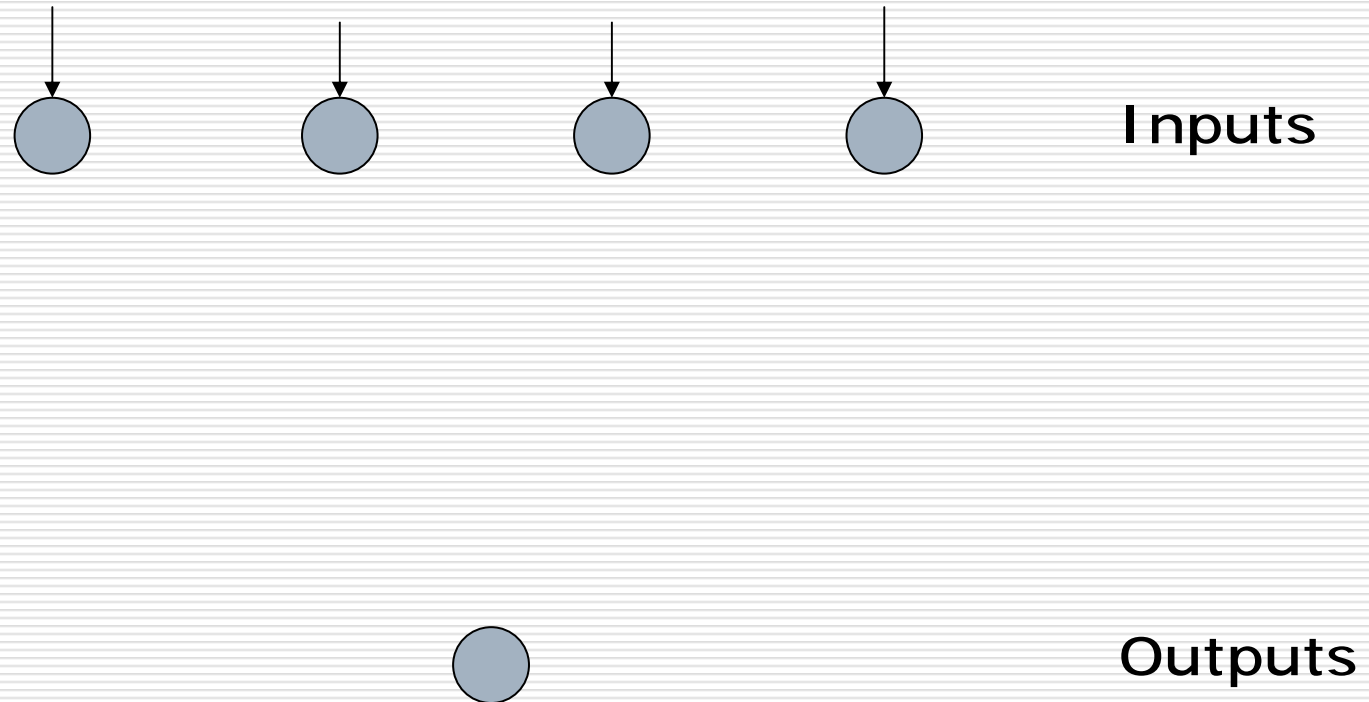
Redes neurales

- ❑ Algoritmo de cálculo que es capaz de aprender la relación entre variables de entrada (predictoras) y salida (sucesos, grupos, etc.)
- ❑ Generalizan cualquier función no-lineal
- ❑ El proceso de aprendizaje es crítico

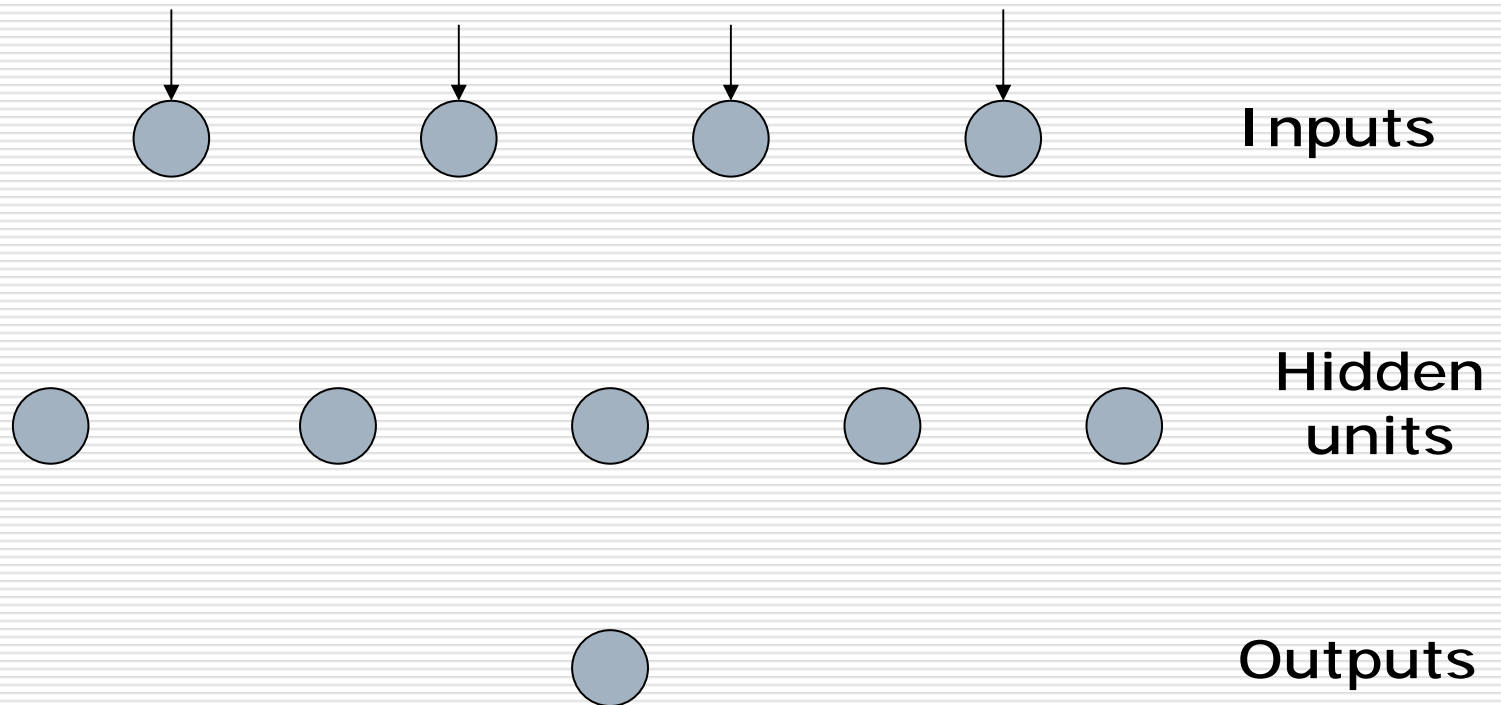
Redes neurales



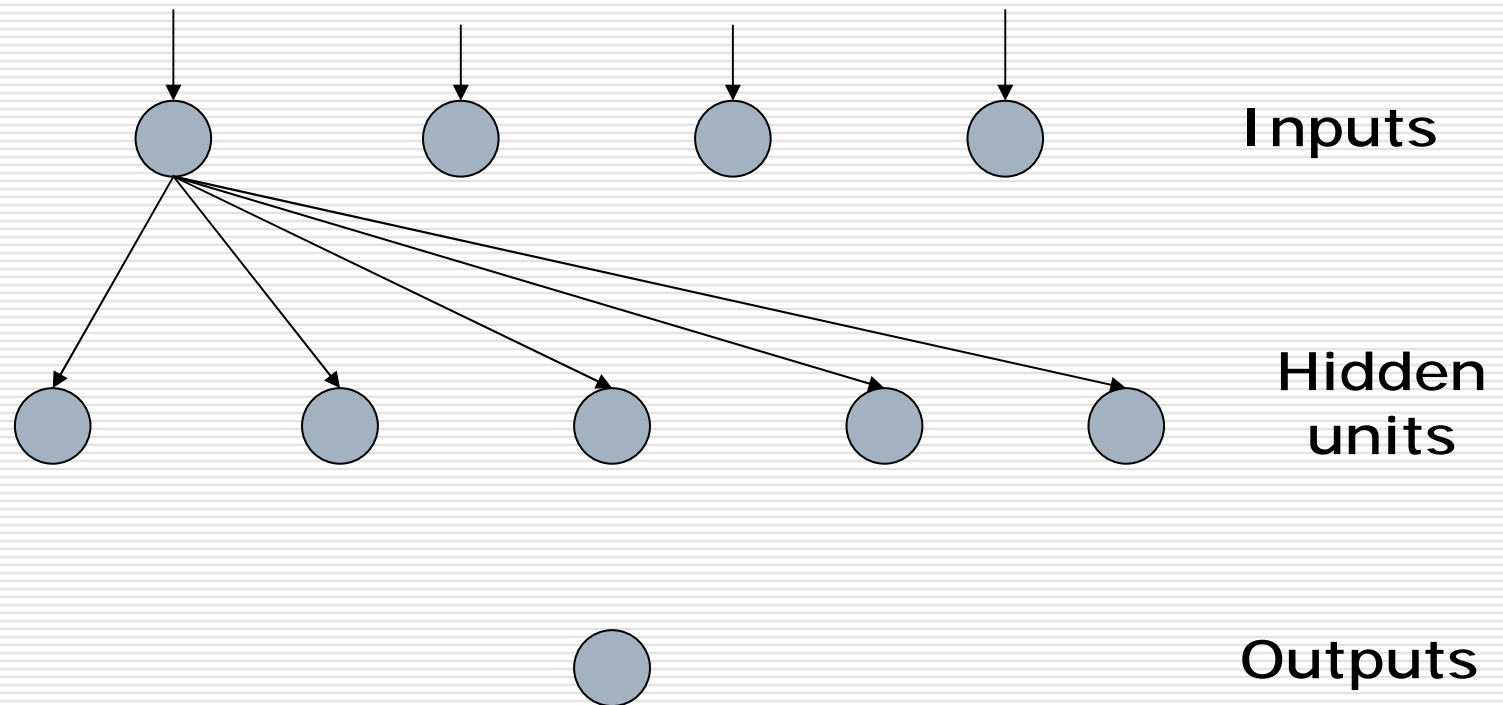
Redes neurales



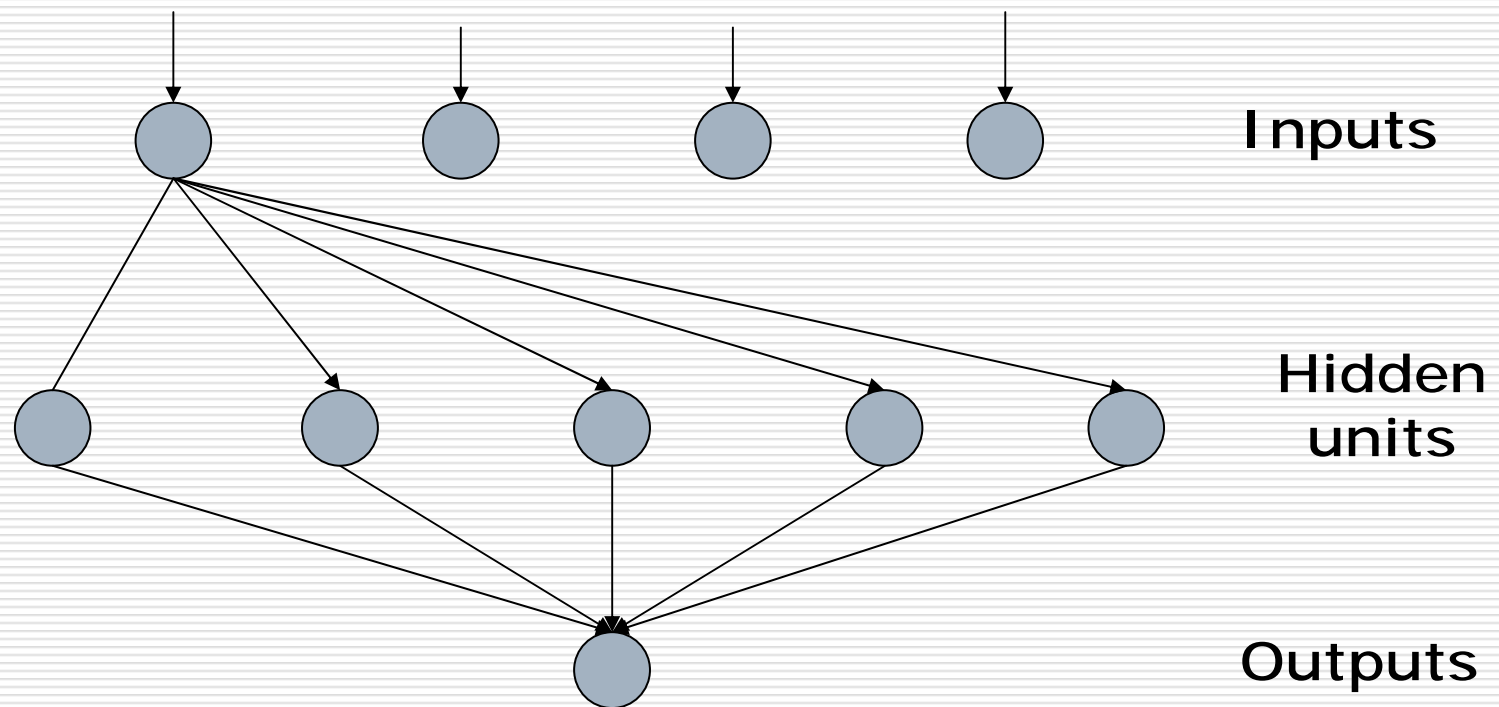
Redes neurales



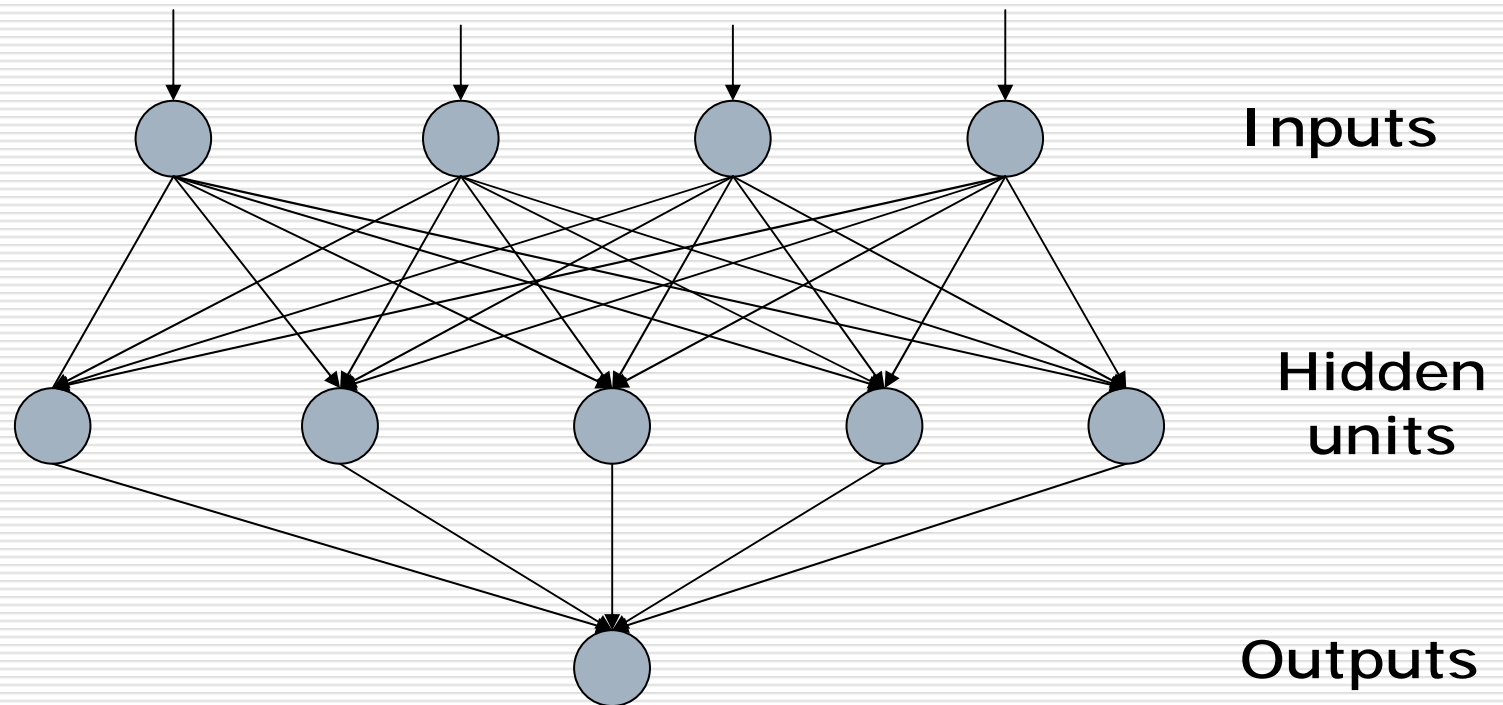
Redes neurales



Redes neurales



Redes neurales



Redes neurales

