



GENERALIZED LINEAR MODELS (GLM) APPLIED TO THE PREDICTION OF HEALTH EXPENDITURE

Vicent Caballer¹, David Vivas¹, Natividad Guadalajara¹, Alexander Zlotnik², Isabel Barrachina¹

1.Ciegs/INECO. Universitat Politècnica de València

2.Universidad Politécnica de Madrid



2017 Spanish Stata Users Group meeting

Madrid | 19 October

Instituto de Salud Carlos III

- Introducción: Clinical Risk Groups
- Objetivos
- Datos y metodología
- Resultados
- Conclusiones

INTRODUCCIÓN

Envejecimiento de la población con su consecuente aumento de las enfermedades crónicas y la innovación de la tecnología médica = Aumento del gasto sanitario a un ritmo superior al de la economía real (PIB)

Es necesario aumentar la eficiencia en la gestión clínica

Los tradicionales modelos para explicar y predecir el gasto sanitario se han basado variables demográficas. Escasa significación estadística

Los modelos de ajuste por riesgo surgen para medir la morbilidad de la población de forma estandarizada, para así conocer cual será el consumo de recursos asistenciales esperado.

Permite una mejor planificación presupuestaria ya que tienen en cuenta el estado de salud de la población y no únicamente la cápita



INTRODUCCIÓN

Los Clinical Risk Groups (CRG) (Hughes et al., 2004) es un sistema de clasificación de pacientes que asigna a cada individuo un grupo de riesgo clínico mutuamente excluyente en función de su historia clínica electrónica.

Su objetivo es establecer grupos homogéneos de gasto sanitario

Utilidades (Inoriza,2009):

- a) Seguimiento de las tasas de prevalencia de enfermedades crónicas
- b) Comprender los patrones de utilización y consumo de servicios
- c) Desarrollar aplicaciones de ajuste de riesgos y precios
- d) Relacionar consumo, satisfacción y medidas de calidad asistencial desde el punto de vista del paciente



INTRODUCCIÓN

CRG Base	ACRG1	ACRG2	ACRG3	Estados de salud CRG
1076	416	151	38	9

1.Estado sano

2.Enfermedad aguda significativa

3.Enfermedad crónica menor única

4.Enfermedad crónica menor en múltiples sistemas orgánicos

5.Enfermedad dominante o crónica moderada única

6.Enfermedad significativa crónica en múltiples sistemas orgánicos

7.Enfermedad dominante crónica en tres o más sistemas orgánicos

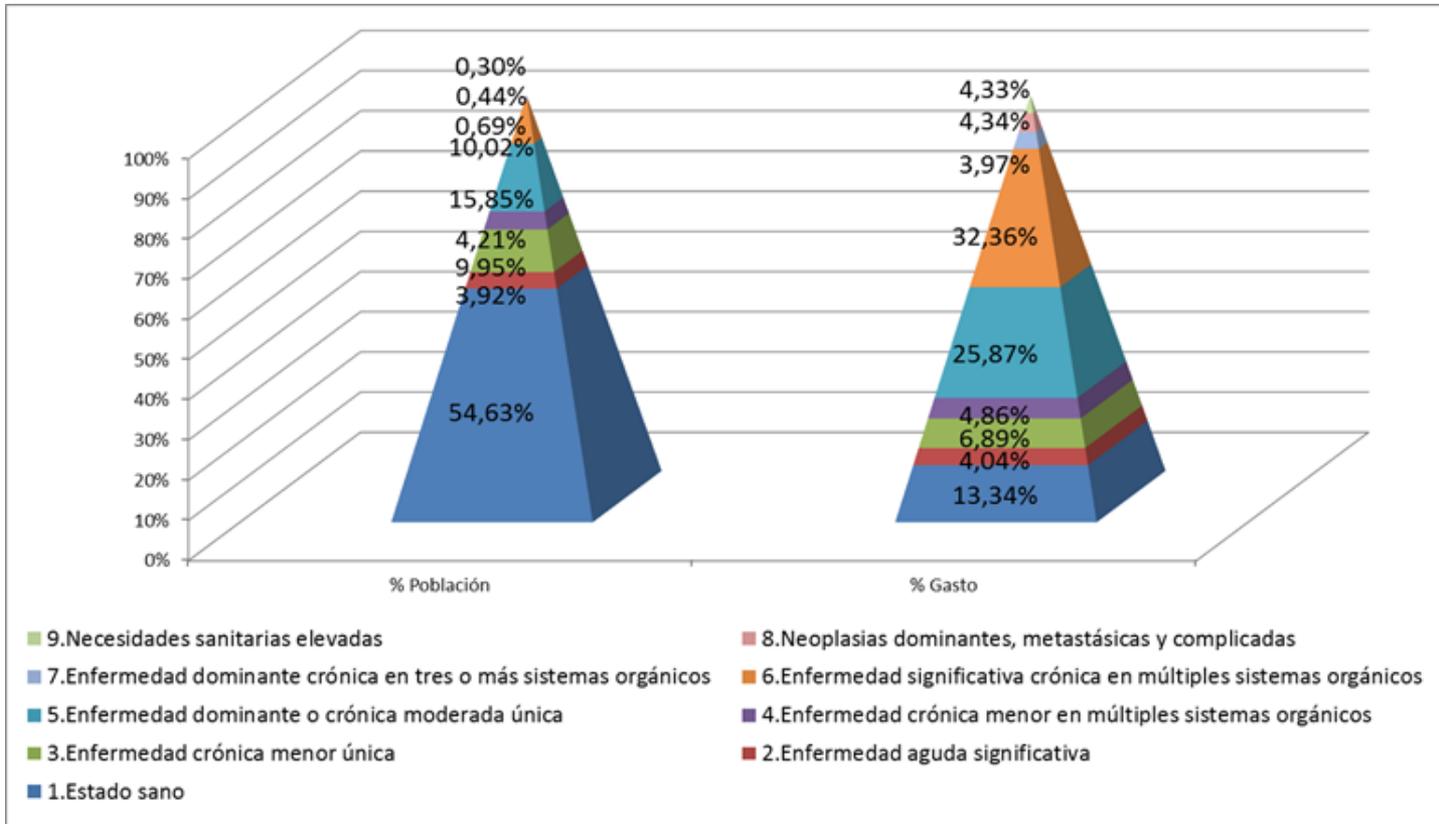
8.Neoplasias dominantes, metastásicas y complicadas

9.Necesidades sanitarias elevadas



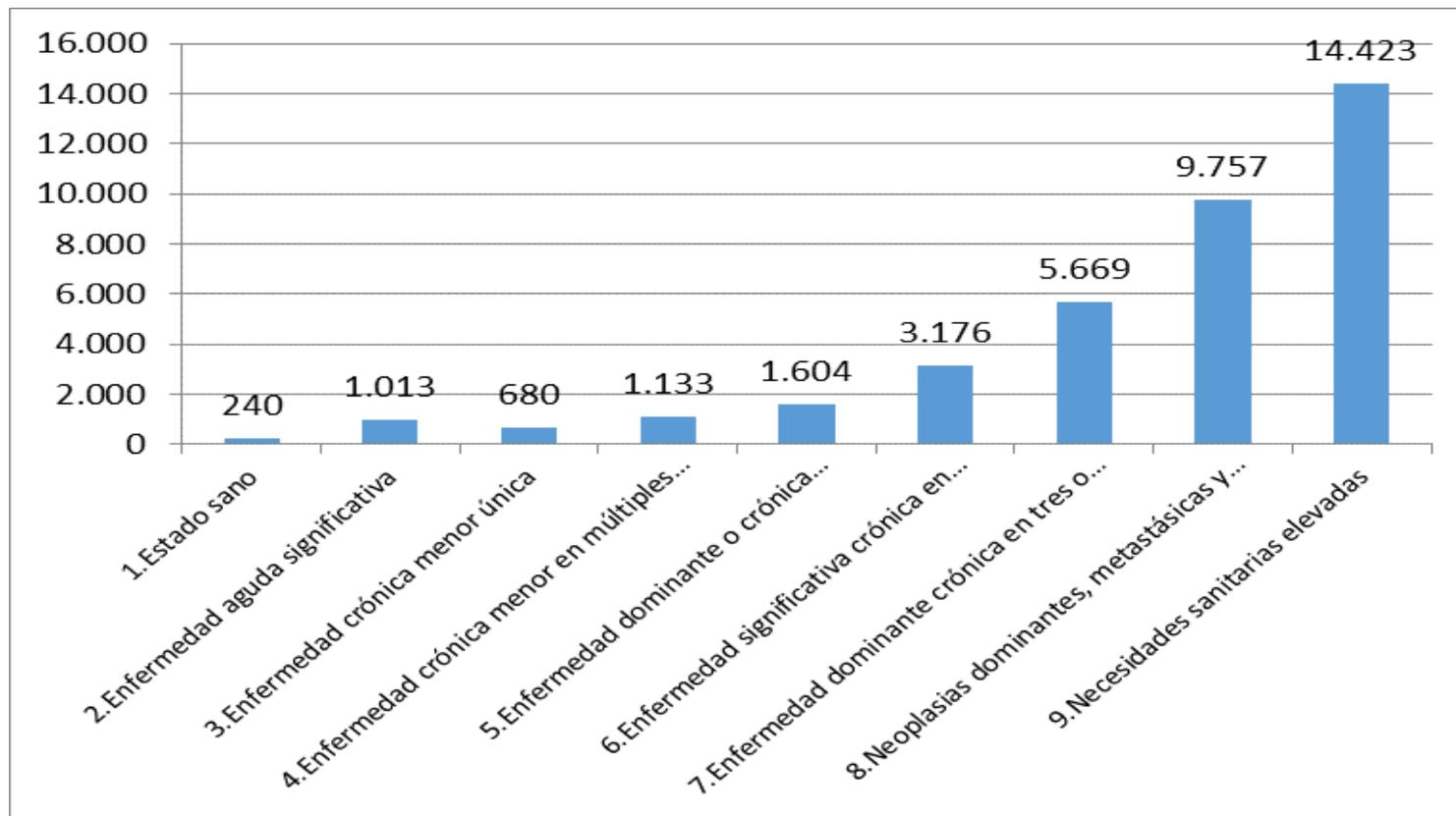
INTRODUCCIÓN

Distribución del gasto sanitario (€) por habitante y de la población por estado de salud en el Departamento de Denia. 2013



INTRODUCCIÓN

Gasto sanitario (€) medio por habitante y estado de salud en el Departamento de Denia



INTRODUCCIÓN

Gasto sanitario (€) total y medio por estado de salud y nivel de gravedad en el Departamento de Denia

Estado de Salud		Nivel de gravedad						TOTAL	
		0	1	2	3	4	5		6
1 Estado sano	Gasto Sanitario	20.560.803						20.560.803	
	Población	85.668						85.668	
	Media	240,01						240,01	
2 Enfermedad aguda significativa	Gasto Sanitario	6.222.537						6.222.537	
	Población	6.142						6.142	
	Media	1.013,11						1.013,11	
3 Enfermedad crónica menor en múltiples sistemas orgánicos	Gasto Sanitario	9.370.832	1.246.942					10.617.774	
	Población	14.805	805					15.610	
	Media	632,95	1.549,00					680,19	
4 Enfermedad crónica menor en múltiples sistemas orgánicos	Gasto Sanitario	3.809.336	1.835.676	1.610.372	229.838			7.485.222	
	Población	4.088	1.425	988	106			6.607	
	Media	931,83	1.288,19	1.629,93	2.168,29			1.132,92	
5 Enfermedad dominante o crónica moderada única	Gasto Sanitario	24.982.953	8.948.721	3.949.288	879.551	1.033.506	71.547	39.865.567	
	Población	18.364	4.495	1.483	183	313	12	24.850	
	Media	1.360,43	1.990,82	2.663,04	4.806,29	3.301,94	5.962,24	1.604,25	
6 Enfermedad significativa crónica en múltiples sistemas orgánicos	Gasto Sanitario	19.267.750	12.508.098	8.937.757	5.919.641	2.862.799	379.055	49.875.100	
	Población	8.244	3.606	2.137	1.202	480	37	15.706	
	Media	2.337,18	3.468,69	4.182,39	4.924,83	5.964,16	10.244,73	3.175,54	
7 Enfermedad dominante crónica en tres o más sistemas orgánicos	Gasto Sanitario	1.137.179	1.065.017	2.640.730	779.800	381.437	112.301	6.116.463	
	Población	292	219	423	96	39	10	1.079	
	Media	3.894,45	4.863,09	6.242,86	8.122,91	9.780,43	11.230,08	5.668,64	
8 Neoplasias dominantes, metastásicas y complicadas	Gasto Sanitario	763.859	2.404.672	2.360.788	929.477	234.478			6.693.274
	Población	107	249	226	81	23			686
	Media	7.138,87	9.657,32	10.445,96	11.475,03	10.194,70			9.756,96
9 Necesidades sanitarias elevadas	Gasto Sanitario	245.664	1.731.729	2.245.323	1.240.186	829.240	385.927	6.678.068	
	Población	59	221	87	59	28	9	463	
	Media	4.163,80	7.835,88	25.808,31	21.020,09	29.615,70	42.880,79	14.423,47	
TOTAL	Gasto Sanitario	26.783.340	59.577.574	29.740.855	21.744.257	9.978.493	5.341.459	948.830	154.114.808
	Población	91.810	45.959	11.020	5.344	1.727	883	68	156.811
	Media	291,73	1.296,32	2.698,81	4.068,91	5.777,93	6.049,22	13.953,38	982,81



OBJETIVOS

Modelizar el gasto sanitario total de los residentes de un área de salud aplicando dos metodologías estadísticas:

Regresión lineal por mínimos cuadrados ordinarios (RL_MCO)

Modelos lineales generalizados (GLM)

Variables explicativas: Variables demográficas y Grupos CRG y niveles de gravedad



DATOS Y METODOLOGÍA

Datos:

Se recogieron datos de un total de **156.811** habitantes del Departamento de Salud de Denia de enero a diciembre de 2013.

Variables:

Se ha diseñado una base de datos que contenía el coste sanitario por individuo, incluyendo los gastos farmacéuticos, hospitalarios y de Atención Primaria en el Departamento de Salud de Denia (Alicante) y el grupo CRG.

Modelización:

Se aplicaron tanto la regresión lineal logarítmica por mínimo cuadrados como el modelo lineal generalizado; la variable dependiente fue el gasto en salud y las variables independientes: el estado de salud del CRG y las variables demográficas (edad y sexo)

Aplicación informática:

Stata 14



DATOS Y METODOLOGÍA

-

Regresión lineal por mínimos cuadrados ordinarios (MCO)

$$\ln(Y + 1) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon$$

Donde:

- Y : Gasto sanitario por paciente en el periodo t .
- X_1, X_2, \dots, X_n : Variables independientes
- $\beta_1, \beta_2, \dots, \beta_n$: parámetros que miden la influencia de la variable X_i sobre el gasto sanitario Y .
- ε : residuo

Cumplimiento de hipótesis (Homocedasticidad, incorrelación, linealidad, normalidad, independencia entre variables).

Retransformación logarítmica de la regresión lineal por mínimos cuadrados ordinarios (MCO)

Smearing Estimator (Duan,1983)

$$E(Y + 1) = E \left(e^{(\alpha + \sum x_i \beta_i) + \varepsilon} \right) = \frac{1}{n} \sum_{i=1}^n \left(e^{(\alpha + x_i \beta_i) + \varepsilon} \right) = \left(e^{(x_i \beta_i)} \right) \left(\frac{1}{n} \sum_{i=1}^n (e^{\varepsilon}) \right) = \left(e^{(x_i \beta_i)} \right) \gamma$$

Factor corrector por la transformación logarítmica



DATOS Y METODOLOGÍA

Modelos Lineales Generalizados (GLM) (McCullagh, Nelder, 1989)

- $E(Y)=\mu=g^{-1}\left(\sum_{j=1}^p \beta_j X_{ji}\right) = g^{-1} X_i \beta$

Donde:

$E(Y)$: Esperanza de la variable dependiente, es decir, el gasto sanitario total.

g : Función de ligadura canónica. Logarítmica gamma

$X_i \beta_j$: Es el predictor lineal denotado por η y que corresponde al vector de n componentes siendo cada una de ellas igual a $\eta_i = \sum_{j=1}^p \beta_j X_{ji} = X_i \beta$

Los GLM se estima por máxima verosimilitud, esto quiere decir que el valor del parámetro analizado tomará el valor que haga máxima la probabilidad de la muestra observada.

Bondad del ajuste: Criterio de información de Akaike (AIC)

$$AIC = (2k - 2 \ln(L)) / N$$

Donde:

k : es el número de parámetros del modelo

L : es el valor máximo de la función de verosimilitud



DATOS Y METODOLOGÍA

	Ventajas	Inconvenientes
Regresión lineal por mínimos cuadrados ordinarios (MCO)	<ul style="list-style-type: none">• Metodología ampliamente conocida por los profesionales del sector• Fácil interpretación de los indicadores de bondad de ajuste	<ul style="list-style-type: none">• Requiere el cumplimiento de las hipótesis
Modelos lineales Generalizados (GLM)	<ul style="list-style-type: none">• No requiere el cumplimiento de las hipótesis	<ul style="list-style-type: none">• Metodología relativamente nueva, no es muy conocida por los profesionales del sector• El criterio de información de Akaike no aporta información por sí mismo. Necesita ser comparado con otros modelos.



RESULTADOS. REGRESIÓN LINEAL

Variables	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5
	Edad y sexo	Estado de salud CRG	Estado de salud CRG y nivel de gravedad	Estado de salud CRG y edad y sexo	Estado de salud CRG, nivel de gravedad, edad y sexo
Constante	3,481	4,124	3,38	3,74	3,76
Edad	0,444			-0,02	-0,02
Sexo	0,027			0,32	0,32
Estado de salud 2		2,824	2,82	2,84	2,84
Estado de salud 3		2,369	2,34	2,54	2,52
Estado de salud 4		3,199	2,95	3,5	3,25
Estado de salud 5		3,288	3,12	3,61	3,44
Estado de salud 6		4,22	3,88	4,74	4,39
Estado de salud 7		4,906	4,36	5,56	4,98
Estado de salud 8		4,991	4,39	5,45	4,81
Estado de salud 9		5,631	5,01	5,95	5,28
Nivel de gravedad 2			0,54		0,58
Nivel de gravedad 3			0,76		0,84
Nivel de gravedad 4			0,96		1,09
Nivel de gravedad 5			1,14		1,31
Nivel de gravedad 6			1,57		1,67
F	5161,941	12966,873	8141,803	11044,669	7544,338
R ²	0,062	0,398	0,403	0,413	0,419



RESULTADOS. MODELOS LINEALES GENERALIZADOS

Variables	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5
	Edad y sexo	Estado de salud CRG	Estado de salud CRG y nivel de gravedad	Estado de salud CRG y edad y sexo	Estado de salud CRG, nivel de gravedad, edad y sexo
Constante	5,74	5,48	5,48	5,54	5,55
Edad	0,02			-0,01	-0,01
Sexo	0,05			0,20	0,21
Estado de salud 2		1,44	1,44	1,45	1,45
Estado de salud 3		1,04	1	1,13	1,09
Estado de salud 4		1,55	1,32	1,68	1,45
Estado de salud 5		1,9	1,73	2,06	1,89
Estado de salud 6		2,58	2,25	2,8	2,47
Estado de salud 7		3,16	2,67	3,43	2,94
Estado de salud 8		3,7	3,2	3,9	3,4
Estado de salud 9		4,09	3,45	4,3	3,63
Nivel de gravedad 2			0,43		0,43
Nivel de gravedad 3			0,63		0,65
Nivel de gravedad 4			0,85		0,88
Nivel de gravedad 5			0,95		1
Nivel de gravedad 6			1,49		1,49
Criterio de información de Akaike (AIC)	15,47929	14,63199	14,23991	14,61018	14,22108



RESULTADOS. MODELOS LINEALES GENERALIZADOS

Comparación de las metodologías. Correlación entre el valor estimado y el real

	Regresión log-MCO	GLM
\bar{R}^2	24,9%	25,7%



El modelo más funcional para predecir el gasto sanitario es el GLM que incluye como variables explicativas los grupos CRG y nivel de gravedad.

CONCLUSIONES. MODELOS LINEALES GENERALIZADOS

A partir del modelo seleccionado se pueden desarrollar diversas aplicaciones:

1. Desarrollo de un modelo de Case Mix
2. Calcular el coste de los pacientes crónicos complejos considerando la multimorbilidad
3. Mejorar la planificación presupuestaria en las diferentes áreas de salud conociendo el estado de salud de la población
4. Predecir la frecuentación de los servicios asistenciales



MUCHAS GRACIAS

dvivas@upv.es

Referencias

Caballer Tarazona, V., Guadalajara Olmeda, N., Vivas Consuelo, D. & Clemente Collado, A. [Impact of Morbidity on Health Care Costs of a Department of Health through Clinical Risk Groups. Valencian Community, Spain]. *Rev. Esp. Salud Publica* **90**, e1–e15 (2016).
