

ESTADÍSTICA MÍNIMA PARA EL REUMATÓLOGO

L. SILVA FERNÁNDEZ

CONTENIDO

INTRODUCCIÓN

RIESGO RELATIVO

RAZÓN DE MOMIOS (*ODDS RATIO*)

INTERVALO DE CONFIANZA DEL 95%

Intervalo de confianza del 95% para la media

Intervalo de confianza del 95% para una proporción

SENSIBILIDAD Y ESPECIFICIDAD

CURVAS ROC

ANÁLISIS UNIVARIANTE, BIVARIANTE Y MULTIVARIANTE

VALOR DE P Y SIGNIFICACIÓN ESTADÍSTICA

BIBLIOGRAFÍA RECOMENDADA

INTRODUCCIÓN

La estadística es la ciencia que describe y analiza datos numéricos individuales y, a partir de ellos, elabora unas conclusiones generales. Puede utilizarse en numerosas áreas de la actividad humana y, aplicada a las ciencias de la vida (entre ellas la medicina), se denomina bioestadística. La estadística, junto con la epidemiología, es un instrumento indispensable en el proceso de investigación en medicina. Formalmente, la estadística se puede clasificar en descriptiva, cuando se utiliza simplemente para la presentación y síntesis de la información recogida en un estudio, e inferencial, que tiene por objetivo generalizar la información obtenida en una muestra a resultados válidos para la población de la que procede.

En este capítulo se presentan conceptos básicos de estadística que ayudarán al reumatólogo clínico a la interpretación de los resultados de artículos científicos y al análisis estadístico básico de datos.

RIESGO RELATIVO

El riesgo constituye una medida de probabilidad estadística de que en un futuro se produzca un acontecimiento, por lo general, no deseado. Implica, por tanto, que la presencia de una característica o factor aumenta la probabilidad de consecuencias adversas. Un factor de riesgo es cualquier característica o circunstancia detectable de una persona o grupo de personas que se sabe asociada con un aumento en la probabi-

lidad de padecer, desarrollar o estar especialmente expuesto a una enfermedad. Estos factores de riesgo, sumándose unos a otros, pueden aumentar el efecto aislado de cada uno de ellos produciendo un fenómeno de interacción.

Hay diferentes maneras de cuantificar el riesgo. Entre ellas, el riesgo absoluto mide la incidencia del evento o fenómeno en la población total y se calcula dividiendo el número de individuos que presentan el evento entre el número total de individuos susceptibles de presentarlo en la población. El riesgo relativo (RR) compara la frecuencia con que ocurre el daño entre los que tienen el factor de riesgo y los que no lo tienen. El RR mide la fuerza de la asociación entre la exposición y la enfermedad. Indica la probabilidad de que se desarrolle la enfermedad en los expuestos a un factor de riesgo en relación con el grupo de los no expuestos y su cálculo se estima dividiendo la incidencia de la enfermedad en los expuestos entre la incidencia de la enfermedad en los no expuestos. Así, de acuerdo con la tabla 5.1:

$$\text{Riesgo relativo} = \frac{\text{Incidencia en expuestos}}{\text{Incidencia en no expuestos}} = \frac{a/(a+b)}{c/(c+d)}$$

TABLA 5-1

Tabla de contingencia para la estimación del riesgo

	Enfermos	Sanos	Total
Expuestos	a	b	a + b
No expuestos	c	d	c + d
Total	a + c	b + d	a + b + c + d

El RR no tiene dimensiones y su rango se sitúa entre 0 e ∞ . Un RR = 1 indica que no hay asociación entre la presencia del factor de riesgo y el evento. Un RR > 1 indica que la asociación es positiva, es decir, la presencia del factor se asocia a mayor ocurrencia del evento; por el contrario, un RR < 1 indica una asociación negativa, es decir, que el factor sería protector.

El concepto de RR es más difícil de interpretar que el de riesgo absoluto, y hay que tener en cuenta que cuando se habla de enfermedades, un RR alto en una enfermedad rara no implica que el riesgo absoluto sea alto.

Ejemplo: Supongamos que queremos determinar el RR de padecer artrosis de rodilla (definida como un grado radiográfico Kellgren-Lawrence ≥ 2) en sujetos obesos (índice de masa corporal ≥ 25) en la muestra de pacientes definida en la tabla 5.2. El riesgo de padecer artrosis en la población expuesta (obesos) en este ejemplo es $157/1.622 = 0,096$ (9,6%) y en la población no expuesta $81/2.049 = 0,039$ (3,9%). Calculamos el RR dividiendo la incidencia en expuestos entre la incidencia en no expuestos: $0,096/0,039 = 2,5$. Esto significa que un sujeto obeso tiene 2,5 veces el riesgo de un sujeto no obeso de padecer artrosis de rodilla.

RAZÓN DE MOMIOS (ODDS RATIO)

La razón de momios (en inglés, *odds ratio* [OR]) es una medida de asociación entre variables binarias que se usa como estimación del RR en los estudios de casos y controles. La *odds* es el cociente entre la probabilidad de que el episodio de interés ocurra y la probabilidad de que no ocurra y, habitualmente, se estima por el cociente entre el número de veces que ha ocurrido el evento y el número de veces que no ha ocurrido. Mide, por tanto, las veces que es más probable que ocurra el evento que las que no ocurra.

La OR es el cociente entre la *odds* en el grupo con el factor de riesgo o tratamiento y la *odds* en el grupo sin el factor de riesgo o tratamiento. A partir de la tabla 5.1, la OR se calcula según la fórmula:

$$\text{Odds ratio} = \frac{\text{Odds en expuestos}}{\text{Odds en no expuestos}} = \frac{a/c}{b/d} = \frac{a \times d}{b \times c}$$

La OR no tiene dimensiones y tiene un rango de 0 a ∞ . Una OR = 1 indica que no hay asociación entre la presencia de la exposición y el evento. Si la OR > 1, la asociación es positiva, es decir, la presencia de la exposición se asocia a mayor ocurrencia del evento; si la OR < 1, la asociación es negativa, es decir, la presencia de la exposición se asocia a una menor ocurrencia del evento.

La OR se comporta de una manera parecida al RR, aunque su interpretación es menos intuitiva. Si el riesgo es muy bajo, OR y RR son muy parecidos. La OR se puede estimar siempre y de la misma forma mientras que el RR no se puede estimar en los diseños caso-control, ya que no es posible calcular las tasas de incidencia. Mediante regresión logística, la OR se puede ajustar por factores de confusión.

Ejemplo: En la muestra de pacientes definida en la tabla 5.2, la *odds* de padecer artrosis de rodilla en sujetos obesos es $157/1.465 = 0,107$ y en sujetos no obesos $81/1.968 = 0,041$. Calculamos la OR dividiendo la *odds* en sujetos obesos entre la *odds* en sujetos no obesos: $0,107/0,041 = 2,6$. Esto significa que los sujetos obesos tienen 2,6 veces más riesgo de padecer artrosis de rodilla que los no obesos.

INTERVALO DE CONFIANZA DEL 95%

Un estimador puntual proporciona la mejor aproximación posible al valor verdadero del parámetro poblacional que se desea estimar y en realidad es desconocido. Sin embargo, el resultado de la estimación puntual no ha de coincidir forzosamente con el verdadero valor del parámetro. De hecho, en muchas ocasiones se puede estar seguro de la no coincidencia. La estimación por intervalos de confianza (IC) proporciona un rango de valores entre los que se tendrá una cierta certeza o nivel de confianza de que se encuentre el parámetro poblacional desconocido. Por tanto, en el contexto de estimar un parámetro poblacional, un IC es un rango de valores, calculado en una muestra, en el cual se encuentra el verdadero valor del parámetro con una probabilidad determinada. La probabilidad "a priori" de que el verdadero valor del parámetro se encuentre en el intervalo construido se denomina *nivel de confianza* y se expresa como $1 - \alpha$. La probabilidad de equivocarnos se llama *nivel de significancia* y se simboliza con α . Generalmente se construyen IC $1 - \alpha = 95\%$ (o significancia $\alpha = 5\%$). Menos frecuentes son los intervalos con $\alpha = 10\%$ o $\alpha = 1\%$.

Intervalo de confianza del 95% para la media

De una población, en la que el parámetro a estimar tenga una media μ y una desviación típica σ , se pueden tomar varias muestras de n elementos. Cada una de estas muestras tendrá, a su vez, una media \bar{x} . La media de todas estas medias muestrales coincidirá con la media poblacional.

La distribución de \bar{x} tiende a ser normal a medida que el tamaño de la muestra se incrementa. En una distribución normal, fijado un nivel de confianza del 95%, se podrá calcular el IC para la media según la fórmula:

$$\mu \pm 1,96 \frac{\sigma}{\sqrt{n}}$$

Ejemplo: En una muestra de 300 sujetos, la edad media es de 54 años con una desviación típica de 14. El IC del 95% es $54 \pm 1,96(14/\sqrt{300}) = 54 \pm 1,58 \Rightarrow 52,41-55,58$.

Intervalo de confianza del 95% para una proporción

De la misma forma que para la media, sea una población en la que una determinada proporción de individuos p presentan una característica. De esta población se puede extraer una muestra aleatoria simple de tamaño n en la cual una proporción p presenta la característica en cuestión. Asumiendo que

el tamaño muestral sea suficientemente grande y, por tanto, “p” siga una distribución normal, el IC del 95% se calcularía siguiendo la fórmula:

$$p \pm 1,96 \sqrt{\frac{p(1-p)}{n}}$$

Ejemplo: En la muestra definida en la tabla 5.2 la frecuencia (proporción) de sujetos con artrosis de rodilla es $238 \times 100/3.671 = 6,48$. A partir de esta muestra, el IC del 95% calculado para la proporción es 5,68-7,27:

$$0,0648 \pm 1,96 \sqrt{\frac{0,0648(1-0,0648)}{3.671}} = 0,0648 \pm 0,00796 \Rightarrow (0,0568-0,0727)$$

SENSIBILIDAD Y ESPECIFICIDAD

Generalmente, la exactitud diagnóstica de una prueba se expresa como sensibilidad y especificidad diagnósticas. Cuando se utiliza una prueba dicotómica (una cuyos resultados se puedan interpretar directamente como positivos o negativos), la sensibilidad es la probabilidad de clasificar correctamente a un individuo enfermo, es decir, la probabilidad de que para un sujeto enfermo se obtenga en la prueba un resultado positivo. La sensibilidad es, por tanto, la capacidad del test para detectar la enfermedad. La especificidad es la probabilidad de clasificar correctamente a un sujeto sano, es decir, la probabilidad de que para un sujeto sano se obtenga un resultado negativo. En otras palabras, se puede definir la especificidad como la capacidad para detectar a los sujetos sanos.

Cuando los datos de una muestra de pacientes se clasifican en una tabla de contingencia por el resultado de la prueba y su estado respecto a la enfermedad (tabla 5.3), la sensibilidad se puede estimar como la proporción de sujetos enfermos que obtuvieron un resultado positivo en la prueba diagnóstica. Es decir,

$$\text{Sensibilidad} = \frac{\text{Verdaderos positivos}}{\text{Verdaderos positivos} + \text{falsos negativos}}$$

de ahí que la sensibilidad también se denomine fracción de verdaderos positivos.

En esta misma tabla, la especificidad se calcularía como la proporción de sujetos sanos que obtuvieron un resultado negativo en la prueba diagnóstica. Es decir,

$$\text{Especificidad} = \frac{\text{Verdaderos negativos}}{\text{Verdaderos negativos} + \text{falsos positivos}}$$

de ahí que la especificidad también se denomine fracción de verdaderos negativos.

Ejemplo: Queremos calcular la sensibilidad y especificidad de la ecografía de hombro para el diagnóstico de roturas parciales del tendón del supraespinoso tomando como patrón oro la resonancia magnética (RM), en la muestra definida en la tabla 5.4. Para calcular la sensibilidad dividimos el número de sujetos correctamente diagnosticados de rotura de tendón por la ecografía (verdaderos positivos) entre el número total de

TABLA 5-2

Tabla de contingencia obesidad*artrosis de rodilla

	Kellgren-Lawrence ≥ 2	Kellgren-Lawrence < 2	Total
IMC ≥ 25 kg/m ²	157	1.465	1.622
IMC < 25 kg/m ²	81	1.968	2.049
Total	238	3.433	3.671

IMC: índice de masa corporal.

TABLA 5-3

Tabla de contingencia para el cálculo de sensibilidad y especificidad de una prueba

		Evento	
		+	-
Test	+	VP	FP
	-	FN	VN

FN: falsos negativos; FP: falsos positivos; VN: verdaderos negativos; VP: verdaderos positivos.

TABLA 5-4

Tabla de contingencia ecografía*resonancia magnética (RM) para el diagnóstico de la rotura parcial del tendón supraespinoso

		Rotura parcial del tendón supraespinoso en la RM	
		+	-
Ecografía	+	19	8
	-	3	191

sujetos que realmente tienen rotura según la RM: $19/(19+3) = 19/22 = 0,86$. Para calcular la especificidad dividimos el número de sujetos que no tienen rotura del tendón y en los que la exploración ecográfica ha sido negativa (verdaderos negativos) entre el número total de sujetos que no tienen rotura según la RM: $191/(191+8) = 191/199 = 0,96$. La ecografía en este ejemplo tiene una sensibilidad del 86% y una especificidad del 96% para el diagnóstico de roturas parciales del tendón del supraespinoso.

CURVAS ROC

Como se ha visto en el apartado anterior, la sensibilidad y la especificidad se utilizan para expresar el valor diagnóstico de pruebas con resultados dicotómicos. Cuando se utilizan pruebas cuyos resultados se miden por variables continuas o discretas ordinales, las características operacionales de la prueba cambian según donde se sitúe el punto de corte. Así, por cada punto de corte, se obtendrá un par de valores de sensibilidad y especificidad. Una curva ROC (acrónimo de *receiver operating characteristic*) o curva de rendimiento diagnóstico es una representación gráfica de la sensibilidad frente a “1 – especificidad” para un sistema clasificador binario según se varía el punto de corte. La curva ROC se obtiene representando, para cada punto de corte, la fracción de

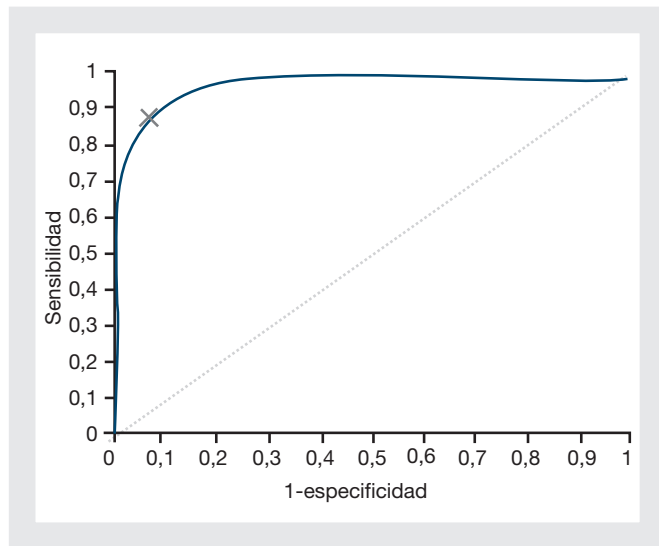


FIGURA 5-1 ■ Curva ROC. La valoración global de la prueba se realiza mediante el área bajo la curva. Esta área posee un valor comprendido entre 0,5 y 1, en donde 1 representa un valor diagnóstico perfecto (la curva pasaría por el vértice superior izquierdo) y 0,5 una prueba sin capacidad discriminativa (la curva sería una recta diagonal).

verdaderos positivos (sensibilidad) en el eje de ordenadas y la fracción de falsos positivos ($1 - \text{especificidad}$) en el eje de abscisas (fig. 5.1). Si la prueba no discriminase entre sanos y enfermos, la curva ROC sería la diagonal que une los vértices inferior izquierdo y superior derecho. La exactitud de la prueba aumenta a medida que la curva se desplaza desde la diagonal hacia el vértice superior izquierdo. Si la discriminación fuera perfecta (100% de sensibilidad y 100% de especificidad) pasaría por dicho punto.

Las curvas ROC son una medida global de la utilidad diagnóstica y un instrumento muy útil para seleccionar el punto de corte óptimo para un determinado test diagnóstico. La valoración global de la prueba se realiza mediante el área bajo la curva. Esta área posee un valor comprendido entre 0,5 y 1, en donde 1 representa un valor diagnóstico perfecto y 0,5 una prueba sin capacidad discriminativa. El área bajo la curva representa el porcentaje de sujetos bien clasificados como sanos o enfermos por la prueba.

ANÁLISIS UNIVARIANTE, BIVARIANTE Y MULTIVARIANTE

El análisis estadístico se puede organizar en 3 fases: el análisis descriptivo (univariante), el análisis bivariante y el análisis multivariante.

El análisis univariante está basado en una sola variable. La naturaleza de la variable a analizar (nominal, ordinal o cuantitativa) determinará las medidas más adecuadas para su análisis. Para las variables nominales, el análisis se limita a las frecuencias por cada categoría. Para las variables ordinales, la tendencia central se puede medir con la moda y la mediana, y

para la dispersión se puede utilizar un diagrama de frecuencias. En cuanto a las variables cuantitativas se puede utilizar la media o mediana para la tendencia central y la desviación típica, rango intercuartílico y varianza para la dispersión.

Las técnicas estadísticas bivariantes permiten el análisis conjunto de 2 características de los sujetos de una población con el propósito de detectar posibles relaciones entre ellas. Se usan para describir una variable (dependiente) en función de la otra (independiente). La herramienta estadística utilizada dependerá también del tipo de variable. En la tabla 5.5 se muestran los tests estadísticos más adecuados para examinar la presencia de asociación entre 2 variables dependiendo de su tipo. Medidas como el RR, la OR o el IC dan información sobre la magnitud de esa asociación.

El análisis multivariante se utiliza cuando se tienen varias variables de cada sujeto en 1 o más muestras. En la biología y en la clínica habitual estamos acostumbrados a ver que diferentes factores causales actúan juntos para producir una consecuencia. El análisis multivariante permite establecer la responsabilidad que tienen diferentes variables predictoras en el resultado clínico que nos interesa, ajustando o controlando por la influencia de cada una de ellas (variables confusoras). Los tests estadísticos a utilizar dependerán una vez más de las variables que queramos analizar (tabla 5.6). Las técnicas multivariantes más utilizadas actualmente en la investigación clínica son las de regresión (lineal, logística y riesgo proporcional de Cox).

La regresión logística multivariante se utiliza cuando la variable dependiente es categórica. La naturaleza de las variables predictoras puede ser categórica, dicotómica o cuantitativa. La regresión lineal múltiple se utiliza cuando la variable dependiente es cuantitativa. En ambos tipos de regresión, las variables independientes se pueden introducir en el modelo de diferentes maneras:

- Introducir todas las variables juntas en un solo paso.
- *Forward* (hacia delante). Introducir progresivamente en el modelo las variables estadísticamente significativas en el análisis bivariante, comenzando por la que tenga mayor significación.
- *Backward* (hacia atrás). Comienza por estimar un modelo que contiene todas las variables predictoras para ir eliminando posteriormente una a una, de forma sucesiva, las que no sean significativas.
- *Stepwise*. Es una combinación de las anteriores. Las variables se van introduciendo secuencialmente según su significación pero, tras incluir cada una de ellas, se comprueba el modelo y se quitan aquellas otras variables que hayan dejado de ser significativas.

La regresión de Cox se utiliza para el análisis de supervivencia. En ese tipo de regresión, el interés es buscar variables independientes que se relacionen con variaciones en la función de supervivencia, o en la función de riesgo, de unos sujetos respecto a un determinado suceso estudiado. Su instrumento fundamental es la *hazard ratio*.

TABLA 5-5
Métodos para el análisis bivariante

Variable independiente	Variable dependiente			
	Dicotómica	Catagórica (> 2 categorías)	Ordinal	Cuantitativa
Dicotómica	χ^2 Z de comparación de proporciones Prueba exacta de Fisher (si $n < 25$) McNemar (si medidas pareadas)	χ^2 Q de Cochran (si medidas pareadas)	U de Mann-Whitney Prueba de los rangos de Wilcoxon (si medidas pareadas)	t de Student t de Student para datos pareados
Catagórica (> 2 categorías)	χ^2 Q de Cochran (si medidas pareadas)	χ^2 Q de Cochran (si medidas pareadas)	Kruskal-Wallis Friedman (si medidas pareadas)	Análisis de la varianza Análisis de la varianza de 2 vías (si medidas pareadas)
Ordinal	U de Mann-Whitney Wilcoxon (si medidas pareadas)	Kruskal-Wallis Friedman (si medidas pareadas)	Correlación de Spearman Regresión de Passing-Bablok	Correlación de Spearman
Cuantitativa	t de Student	Análisis de la varianza	Correlación de Spearman Tau de Kendall	Correlación de Pearson Correlación de Spearman

TABLA 5-6
Métodos para el análisis multivariante

Variable independiente	Variable dependiente				Supervivencia
	Dicotómica	Catagórica (> 2 categorías)	Ordinal	Cuantitativa	
Dicotómica	Regresión logística multivariante Análisis discriminante Análisis conjunto Análisis de correspondencias	Regresión logística condicional Análisis discriminante Análisis conjunto Análisis de correspondencias	Regresión lineal Regresión de Poisson Análisis discriminante Análisis conjunto	Regresión lineal múltiple ANOVA MANOVA Análisis conjunto	Regresión de Cox
Catagórica (> 2 categorías)	Regresión logística multivariante Análisis discriminante Análisis conjunto Análisis de correspondencias	Regresión logística condicional Análisis discriminante Análisis conjunto Análisis de correspondencias	Regresión lineal Regresión de Poisson Análisis discriminante Análisis conjunto	Regresión lineal múltiple ANOVA MANOVA Análisis conjunto	Regresión de Cox
Ordinal	Regresión logística multivariante Análisis discriminante Análisis conjunto	Regresión logística condicional Análisis discriminante Análisis conjunto	Regresión lineal Regresión de Poisson Análisis discriminante Análisis conjunto	Regresión lineal múltiple ANOVA MANOVA Análisis conjunto	Regresión de Cox
Cuantitativa	Regresión logística multivariante Análisis discriminante	Regresión logística condicional Análisis discriminante	Regresión lineal Regresión de Poisson Análisis discriminante	Regresión lineal múltiple ANCOVA Correlación canónica Análisis factorial y por componentes principales	Regresión de Cox

ANOVA: análisis de la varianza; ANCOVA: análisis de covarianza; MANOVA: análisis multivariante de la varianza.

VALOR DE P Y SIGNIFICACIÓN ESTADÍSTICA

La realización de cualquier estudio pretende determinar si hay o no asociación entre diferentes variables. El hallazgo de una asociación puede deberse a que realmente exista esa asociación o ser producto del azar.

Una hipótesis estadística es una asunción relativa a una o varias poblaciones, que puede ser cierta o no. Las hipótesis estadísticas se pueden contrastar con la información extraída de las muestras y, tanto si se aceptan como si se rechazan, se puede cometer un error.

En el contraste de hipótesis, la hipótesis nula (H_0) equivale a la no asociación entre las 2 variables estudiadas y la hipótesis alternativa (H_a) se corresponde con la existencia de algún grado de asociación entre las 2 variables. El valor p (o simplemente, la p) se define como la probabilidad de rechazar la H_0 siendo esta cierta. En otras palabras, es el riesgo de aceptar que hay una asociación o diferencia cuando en realidad no existe. Habitualmente, el valor p escogido para indicar que una asociación es estadísticamente significativa es 0,05. Esto implica una seguridad del 95% para rechazar la H_0 . Equivale a aceptar que hay una asociación o diferencia y que solo hay un 5% de probabilidad de que hubiésemos obtenido este resultado como producto del azar. Es importante recalcar que un contraste de hipótesis no permite aceptar una H_0 ; simplemente la rechaza o no la rechaza. El no rechazarla no implica que la H_0 sea cierta.

La significación estadística de una diferencia o asociación depende fundamentalmente de 2 factores: la magnitud de la diferencia a testar y el tamaño muestral. Cuanto más grande es la diferencia, más fácil es demostrar que es significativa. Grandes tamaños muestrales permiten detectar pequeñas diferencias, por lo que cualquier diferencia puede ser estadísticamente significativa si se dispone del suficiente número de pacientes.

Se debe tener en cuenta que la significación estadística es un concepto matemático, por lo que una asociación estadísticamente significativa puede no ser clínicamente relevante. De la misma manera, que una asociación sea estadísticamente significativa no implica que la relación sea causal y una asociación estadísticamente no significativa puede deberse a un problema de tamaño muestral insuficiente y ser, sin embargo, de gran relevancia clínica.

BIBLIOGRAFÍA RECOMENDADA

- Díaz Portillo J. Guía práctica del curso de bioestadística aplicada a las ciencias de la salud. Madrid, Instituto Nacional de Gestión Sanitaria, 2011.
- Fisterra. Metodología de la Investigación [consultado 11-2013]. Disponible en: <http://www.fisterra.com/formacion/metodologia-investigacion>
- Glantz S. Bioestadística, 6.ª ed. México, McGraw-Hill, 2006.
- Greenhalgh T. How to read a paper: Statistics for the non-statistician. I: Different types of data need different statistical tests. *BMJ* 1997; 315: 364-6.
- Martínez-González MA, Sánchez-Villegas A. Bioestadística amigable, 2.ª ed. Barcelona, Díaz de Santos, 2006.