

# Bioestadística

## Práctica 2

En esta práctica vamos a aprender a analizar (con *Epidat*) los datos provenientes de distintos tipos de estudios epidemiológicos. En cada caso vamos a identificar en la salida que proporciona el programa los distintas clases de medidas vistas en teoría, y reproduciremos algunas de las medidas con la calculadora. Haremos también un ejercicio concreto para estudiar las consecuencias que el sesgo de autoselección tiene en la exactitud de los resultados.

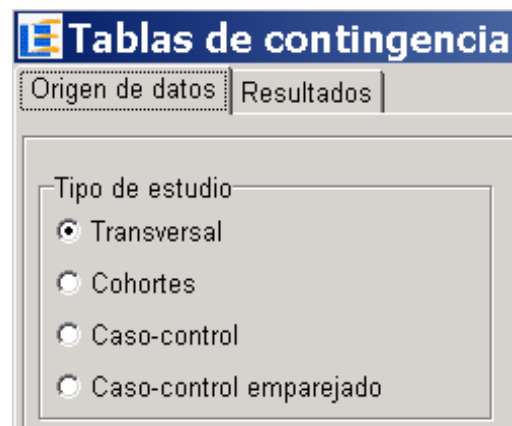
### Introducción

Después de ver un resumen histórico de la Epidemiología, se han definido los conceptos básicos —pero cruciales— en que se apoya esta disciplina: relación de causalidad, diferencia entre asociación estadística y causalidad, características de las relaciones causales, tipos de causas y factor de riesgo. Recordemos que *la existencia de una asociación estadísticamente significativa entre la causa y su efecto es uno de los criterios para proponer una relación causal; aunque hay que tener en cuenta que no es el único.*

En el tema 2 se han clasificado los estudios epidemiológicos según distintos criterios. El programa permite analizar cómodamente los resultados cuando se encuentran ya en tablas de contingencia. Los menús del programa hacen una primera clasificación de los análisis por el tamaño de las tablas y por la existencia o no de estratos en la población del estudio



Una vez seleccionada la forma de la tabla en que se encuentran nuestros datos, *Epidat* trabaja con una clasificación de los tipos de estudio que corresponde a la que en los apuntes denomina «según la selección de la población». El programa va a solicitarnos que especifiquemos si la tabla pertenece a un **estudio transversal, de cohortes o de caso-control** (normal o emparejado). Para recordar estos conceptos, vamos a buscar las definiciones en el artículo «Diseño de estudios epidemiológicos»; Mauricio Hernández-Ávila, Francisco Garrido-Latorre y Sergio López-Moreno; salud pública de México; vol. 42, no.2, 2000.



Por otro lado, se han clasificado también las distintas medidas que se utilizan en Epidemiología en **medidas de frecuencia, medidas de asociación y medidas de impacto**. Para recordar estas definiciones, vamos a utilizar el artículo «Principales medidas en epidemiología»; Alejandra Moreno-Altamirano, Sergio López-Moreno y Alexander Corcho-Berdugo; salud pública de México; vol. 42, no. 4, 2000. Recordemos también la relación que hay entre los niveles de actuación de la Epidemiología—descriptivo, de conocimiento etiológico y de intervención— y los tipos de medidas—de frecuencia, de asociación y de impacto—, respectivamente.

En el mismo artículo se explican tanto las principales escalas de medición para esas medidas, **escalas nominal, ordinal, de intervalo y de razón**, como los conceptos matemáticos con que se definen las medidas, **proporción, tasa y razón**. Comprender bien la diferencia entre estos tres conceptos ayuda mucho a saber por qué las medidas se definen como se definen y cómo interpretarlas. Sacamos del artículo las primeras palabras de cada definición:

**Proporciones:** *Las proporciones son medidas que expresan la frecuencia con la que ocurre un evento en relación con la población total en la cual éste puede ocurrir.*

**Tasas:** *Las tasas expresan la dinámica de un suceso en una población a lo largo del tiempo. Se pueden definir como la magnitud del cambio de una variable (enfermedad o muerte) por unidad de cambio de otra (usualmente el tiempo) en relación con el tamaño de la población que se encuentra en riesgo de experimentar el suceso.*

**Razones:** *Las razones pueden definirse como magnitudes que expresan la relación aritmética existente entre dos eventos en una misma población, o un solo evento en dos poblaciones.*

Es interesante ahora que dediquemos unos minutos a pensar en las siguientes cuestiones:

- ¿Entre qué valores estará, por su propia definición, cada uno de estos conceptos?
- ¿Qué unidades tendrán?
- ¿Cuáles serán adecuadas, o tendrán sentido, para cada tipo de estudio?

## Ejercicio 1

Como primer ejercicio de la práctica, y después de esta larga introducción, vamos a estudiar la salida que *Epidat* proporciona para cada tipo de estudio. Además, para poder comparar las distintas salidas, vamos a hacer seguidamente cinco análisis. Los datos están tomados del archivo de ayuda de *Epidat* del tema de tablas de contingencia, donde se explican e interpretan los resultados. Nosotros insistiremos en la interpretación de estos resultados en la práctica siguiente, después de estudiar teóricamente este tipo de estudios.

Para **analizar una tabla 2x2 simple** entramos en el menú:

*Métodos --> Tablas de contingencia --> Tablas 2x2 --> Simples*

Una vez dentro del menú, para analizar datos de un **estudio transversal** introducimos los datos como se indica en la figura siguiente y después pulsamos el botón de cálculo

Tablas de contingencia: Tablas 2x2 simples

Origen de datos | Resultados

Tipo de estudio

- Transversal
- Cohortes
- Caso-control
- Caso-control emparejado

Nivel de confianza (%)

95,0

Sumar 0,5 a todas las frecuencias si hay ceros

		Enfermedad		Total
		Enfermos	Sanos	
Factor de riesgo	Expuestos	58	62	120
	No expuestos	22	258	280
Total		80	320	400

Para analizar datos de un **estudio de cohortes** (utilizando la **incidencia acumulada**) introducimos los datos como en la figura siguiente y pulsamos el botón de cálculo

**Tablas de contingencia: Tablas 2x2 simples**

Origen de datos | Resultados

Tipo de estudio:  
 Transversal  
 Cohortes  
 Caso-control  
 Caso-control emparejado

Tipo de datos:  
 Tasa de incidencia  
 Incidencia acumulada

Nivel de confianza (%)

Sumar 0,5 a todas las frecuencias si hay ceros

Enfermedad

		Enfermos	Sanos	Total
Factor de riesgo	Expuestos	76	6169	6245
	No expuestos	28	7867	7895
Total		104	14036	14140

Obsérvese que en este caso el programa nos ofrece trabajar con una de las dos medidas de frecuencia siguientes: la tasa de incidencia o la incidencia acumulada. Para analizar datos de un **estudio de cohortes** (utilizando la **tasa de incidencia**) introducimos los datos como en la figura siguiente y pulsamos el botón de cálculo

**Tablas de contingencia: Tablas 2x2 simples**

Origen de datos | Resultados

Tipo de estudio:  
 Transversal  
 Cohortes  
 Caso-control  
 Caso-control emparejado

Tipo de datos:  
 Tasa de incidencia  
 Incidencia acumulada

Nivel de confianza (%)

Sumar 0,5 a todas las frecuencias si hay ceros

Factor de riesgo

		Casos	Personas-Tiempo
Factor de riesgo	Expuestos	76	116157
	No expuestos	28	177636
Total		104	293793

Para analizar datos de un **estudio de casos y controles** introducimos los datos como se indica en la figura siguiente y después pulsamos el botón de cálculo

**Tablas de contingencia: Tablas 2x2 simples**

Origen de datos | Resultados

Tipo de estudio

- Transversal
- Cohortes
- Caso-control
- Caso-control emparejado

Nivel de confianza (%)

Sumar 0,5 a todas las frecuencias si hay ceros

Enfermedad

		Casos	Controles	Total
Factor de riesgo	Expuestos	255	487	742
	No expuestos	500	268	768
Total		755	755	1510

Por último, para analizar los datos del **estudio de casos y controles emparejados**

**Tablas de contingencia: Tablas 2x2 simples**

Origen de datos | Resultados

Tipo de estudio

- Transversal
- Cohortes
- Caso-control
- Caso-control emparejado

Nº de controles por caso  Nivel de confianza (%)

Sumar 0,5 a todas las frecuencias si hay ceros

Controles

		Expuestos	No expuestos	Total
Casos	Expuestos	160	95	255
	No expuestos	327	173	500
Total		487	268	755

Una vez que hemos hecho los cuatro análisis, pulsando el botón de vista preliminar, accedemos a la ventana en la que se han guardado los resultados de toda la sesión (no sólo del último análisis, como muestra la pestaña de resultados).

El ejercicio ahora consiste en:

- ➔ Identificar las medidas de la salida de cada uno de los análisis por separado. ¿Da el programa las medidas ordenadas en medidas de frecuencia, de asociación y de impacto? ¿Cuáles son de cada tipo?
- ➔ Comparar entre sí las salidas de los distintos tipos de estudio.

## Ejercicio 2

**Los conceptos involucrados en Epidemiología suelen tener definiciones bastante sencillas.** De hecho puede parecer que algunas de estas definiciones «coinciden», en el sentido de que se calculan de la misma forma a partir de los datos de las celdas de las tablas de contingencia. Las diferencias hay que buscarlas en lo que los valores de las celdas significan según el tipo de estudio del que procedan. En este ejercicio vamos a reproducir los cálculos de algunas de las medidas que proporciona el programa: así por un lado repasamos sus definiciones y por otro vemos que muchos de esos cálculos pueden hacerse facilísimamente con una calculadora, o incluso a mano. La salida del programa para el estudio transversal anterior es la siguiente:

```
[ 1] Tablas de contingencia: Tablas 2x2 simples

Tipo de estudio   : Transversal
Nivel de confianza: 95,0%
```

Tabla	Enfermos	Sanos	Total
Expuestos	58	62	120
No expuestos	22	258	280
Total	80	320	400

Prevalencia de la enfermedad	Estimación	IC (95,0%)	
En expuestos	<b>0,483333</b>	-	-
En no expuestos	<b>0,078571</b>	-	-
Razón de prevalencias (Katz)	<b>6,151515</b>	3,955011	9,567897

Prevalencia de exposición	Estimación		IC (95,0%)	
En enfermos	<b>0,725000</b>	-	-	-
En no enfermos	<b>0,193750</b>	-	-	-
Razón de prevalencias (Katz)	<b>3,741935</b>	2,882081	4,858324	

OR	IC (95,0%)		
<b>10,970674</b>	6,243768	19,276133	(Woolf)
	6,264300	19,204815	(Cornfield)

Prueba Ji-cuadrado de asociación	Estadístico	Valor p
Sin corrección	86,0119	0,0000
Corrección de Yates	83,5007	0,0000

Prueba exacta de Fisher	Valor p
Unilateral	0,0000
Bilateral	0,0000

Este ejercicio consiste en buscar las definiciones de los conceptos de esta salida que están en negrita y, aplicando esa definición, reproducir los cálculos que ha hecho el ordenador con una calculadora o a mano. Un ejercicio un poco más difícil —pero no mucho más— consistiría en hacer lo mismo para las medidas del estudio de cohortes, donde habría que calcular las cantidades persona-tiempo.

30/03/07

David Casado de Lucas: <http://www.est.uc3m.es/dcasado/>  
Departamento de Estadística  
Universidad Carlos III de Madrid

