

Análisis de datos en los estudios epidemiológicos I

Introducción

En este capítulo, de continuación de nuestra serie temática de formación en metodología de la investigación, vamos a centrarnos en el análisis de datos de los denominados estudios epidemiológicos. Dada la amplitud del tema, pensamos que serán necesarios varios capítulos para aproximarnos con alguna profundidad al tema en estudio.

Aunque actualmente cualquier estudio epidemiológico, de cierto rigor, utiliza para el análisis de datos paquetes estadísticos, como el EPIINFO y el SPSS, no por ello debemos de pasar de largo sobre los conceptos estadísticos y epidemiológicos, en los que estos programas se basan. Conceptos conocidos por todos los profesionales de la salud pero probablemente olvidados, ya que, como ocurre con otras muchas áreas de conocimiento, la "cercanía" de la estadística se consigue con la práctica.

El análisis de datos es el final de la investigación, el momento en que vamos a conocer de manera precisa qué ha ocurrido con la información, con los datos obtenidos en nuestro estudio.

Esta etapa de análisis a su vez consta de varias fases, entre ellas la de decidir el tipo de análisis que quiero realizar. Ya se indicó, en los capítulos dedicados al diseño de los estudios epidemiológicos, que en el propio diseño de la investigación se debía indicar el tipo de análisis que se va a realizar. También se indicó que a pesar de la existencia de elementos comunes, cada tipo de estudio por su parte requería un análisis concreto y que, siendo cierto que tanto los estudios descriptivos como los analíticos requerían una descripción detallada de las variables en estudio, los estudios de cohortes casos control y experimentales, que pretenden encontrar asociación o causalidad entre variables, precisan además un abordaje más complejo.

Etapas del análisis epidemiológico.

En cualquier estudio bien diseñado los datos obtenidos en el mismo contienen la información necesaria para conseguir los objetivos planteados, datos que deben ser sometidos a distintas etapas de análisis.

- **Comprobación.**

En primer lugar debemos proceder a comprobar que los datos son correctos, tenemos que analizar que no falten datos, que estén completos, valoraremos su precisión o exactitud y su consistencia.

- **Procesamiento de datos, también denominada sumarización.**

La presentación de los datos es un aspecto esencial del proceso de análisis, no podemos dar los datos tal y como los hemos obtenido, estos datos se deben transformar en resultados reduciendo la extensión de los mismos. La sumarización de los datos facilitará el subsiguiente análisis.

La forma más simple de agrupar los datos es utilizar tablas de frecuencia, proceso conocido como tabulación. Dependiendo del tipo de análisis que deseamos realizar deberemos utilizar tablas de distribución de frecuencias o tablas de contingencia, en las que se tabula la distribución de las observaciones en función de las variables estudiadas.

Agrupación de datos. Distribuciones de frecuencias.

Como ya indicamos anteriormente, los datos en bruto no nos informan de casi nada, por el contrario pueden resultarnos abrumadores. Para que nos resulten útiles necesitamos ordenarlos, agruparlos, etc.

El conjunto de datos obtenidos en un estudio se pueden describir en base a tres elementos esenciales:

- La distribución de los valores obtenidos
- La tendencia y posición de los mismos en cuanto a un valor central
- La variabilidad que presentan.

Las distribuciones de frecuencias

Las tablas de distribución de frecuencia

Realizar una tabla de distribución de frecuencias con los datos obtenidos en una investigación es el primer paso para imponer un cierto orden al conjunto de informaciones. Este proceso supone una sistematización de los valores numéricos obtenidos, así como un recuento del número de veces que se ha obtenido ese valor.

Construir una distribución de frecuencias es un proceso simple ya que consta básicamente de dos componentes, los valores obtenidos (X) y el número de veces o frecuencia con la que se ha obtenido ese valor (f). El único requisito de una tabla de frecuencias es que las clases en las que ordenamos los datos sean mutuamente excluyentes y exhaustivas en su conjunto, es decir un valor no puede estar a la vez incorporado a dos clases distintas.

Observemos lo anterior en el siguiente ejemplo:

Pensemos que estamos realizando un estudio para conocer los valores de glucemia basal de los niños de un colegio de primaria. Tomamos una muestra representativa de 25 niños y niñas obtenemos los siguientes valores:

90,105, 106, 79, 90, 100, 103, 105, 90, 98, 90, 95, 100, 103, 102, 105, 90, 110, 86, 90, 100, 110, 90, 86, 90.

Como podemos observar en esta distribución difícilmente podemos apreciar nada.

Procedemos entonces a ordenar los datos, de menor a mayor, colocando los mismos en una tabla de distribución de frecuencias. A menudo nos interesa no sólo saber el recuento de las frecuencias absolutas sino también de las frecuencias relativas y de los porcentajes.

Así como el sumatorio de todas las frecuencias debe ser igual al número de sujetos estudiados (N), la suma de los porcentajes de cada puntuación deber ser igual al 100%.

Valores de glucemia	Recuento	Frecuencia absoluta	Frecuencia relativa o proporción	Porcentaje
79	I	1	0,04	4%
86	II	2	0,08	8%
90	IIIIIIII	8	0,32	32%
95	I	1	0,04	4%
98	I	1	0,04	4%
100	III	3	0,12	12%
102	I	1	0,04	4%
103	II	2	0,08	8%
105	III	3	0,12	12%
106	I	1	0,04	4%
110	II	2	0,08	8%
		N = 25	1	100%

En esta tabla ya tenemos realizada una primera ordenación de los datos.

No nos vamos a detener en la explicación de todos estos conceptos, puesto que entendemos son conocidos por los profesionales de enfermería, pero si en la de algunos que quizás no se hayan estudiado suficientemente.

La frecuencia acumulada

Como su nombre indica es la acumulación de frecuencias que se han producido en cada clase o categoría de la variable en estudio. Se calcula a partir de la categoría inferior a la que se van sumando el resto de los valores de cada una de las categorías. Lógicamente la frecuencia acumulada de la categoría última debe coincidir con el número total de casos. De forma similar se realiza la proporción acumulada y el porcentaje acumulados.

Vamos a realizarlo a partir del ejemplo anterior.

Valores de glucemia	Frecuencia absoluta	Frecuencia relativa o proporción	Porcentaje	Frecuencia acumulada	Proporción acumulada	Porcentaje acumulado
79	1	0,04	4%	1	0,04	4%
86	2	0,08	8%	3	0,12	12%
90	8	0,32	32%	11	0,44	44%
95	1	0,04	4%	12	0,48	48%
98	1	0,04	4%	13	0,52	52%
100	3	0,12	12%	16	0,64	64%
102	1	0,04	4%	17	0,68	68%
103	2	0,08	8%	19	0,76	76%
105	3	0,12	12%	22	0,88	88%
106	1	0,04	4%	23	0,92	92%
110	2	0,08	8%	25	1	100%
	N = 25	1	100%			

Evidentemente el valor de glucemia es una variable cuantitativa continua, que se convierte en una variable discreta con los instrumentos de medida. Si queremos realizar la tabla de distribución de frecuencias para las variables cuantitativas continuas debemos de proceder de otro modo. Debemos en primer lugar ordenar los datos en base a intervalos, para ello procederemos del siguiente modo:

- Determinar la amplitud de los intervalos. Es decir el intervalo dentro del cual varían los datos de la muestra. Se calcula restando la puntuación máxima de la mínima $A = X_{\max} - X_{\min}$.
- Determinar el número de intervalos que vamos a utilizar. Esto dependerá del tamaño de nuestra muestra. Normalmente se suele utilizar una regla que es la raíz cuadrada del número de sujetos (\sqrt{N}).

- Calcular la amplitud del intervalo. La amplitud se suele representar con la letra i y se obtiene dividiendo la amplitud de la distribución por el número de intervalos:

$$I = \text{Amplitud de la distribución} \div \text{número de intervalos}$$

- Determinar el límite inferior del primer intervalo, es decir, la puntuación más pequeña a partir de la cual comenzamos a contar el resto.

Las representaciones gráficas

Con frecuencia los investigadores prefieren presentar sus datos en forma de gráficos. Para realizar una correcta presentación gráfica tenemos que tener en cuenta el tipo de variables de estudio. En el cuadro siguiente presentamos el tipo de tabulación y el gráfico más adecuado a cada tipo de variable lo cual nos permitirá también recordar algunos conceptos sobre las variables, ya tratados en los temas anteriores.

Tipo de variable	Tabulación	Representación gráfica
Cualitativa	Frecuencia Proporción Porcentaje	Diagramas de barras Diagramas de sectores Pictogramas
Cuasi cuantitativa	Frecuencia Proporción Porcentaje Proporción acumulada Frecuencia acumulada Porcentaje acumulado	Diagramas de barras Diagramas de sectores Pictogramas
Cuantitativa Discreta	Frecuencia Proporción Porcentaje Proporción acumulada Frecuencia acumulada Porcentaje acumulado	Diagramas de barras
Cuantitativa continua	Frecuencia Proporción Porcentaje Proporción acumulada Frecuencia acumulada Porcentaje acumulado	Histogramas Polígonos de frecuencias

Diagramas de barras.

Consiste en un conjunto de barras o rectángulos sobre un eje de coordenadas. La altura de cada barra estará determinada por la frecuencia de la modalidad que representa.

Causas de mortalidad infantil

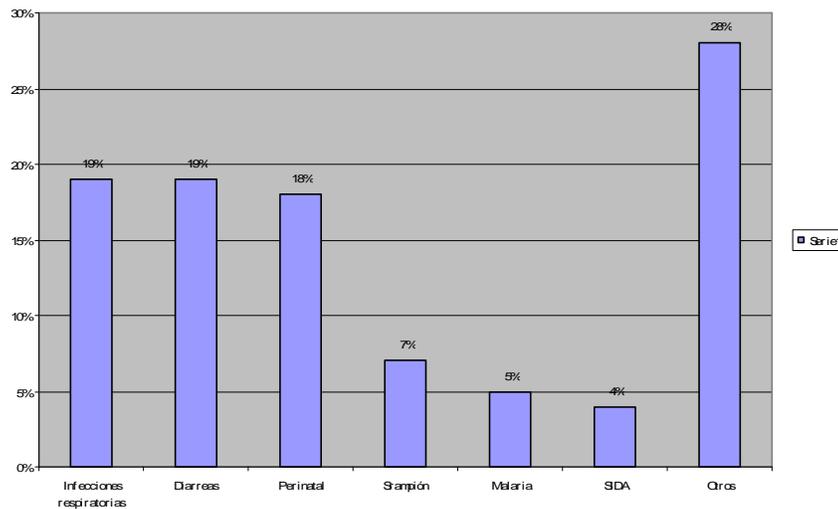
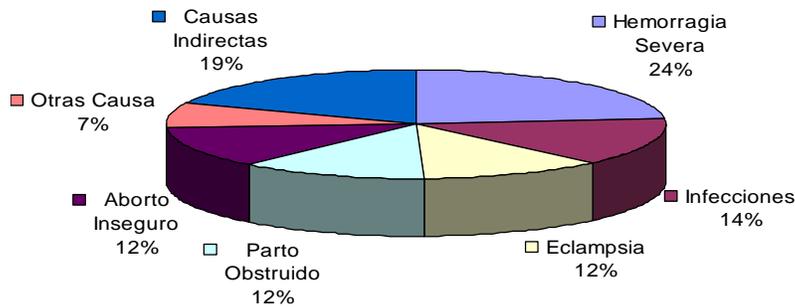


Diagrama de sectores.

Consiste en representar mediante sectores circulares las distintas modalidades de una variable. Cada una de las modalidades se representa proporcionalmente a los 360° del círculo.

Causas de Mortalidad Materna

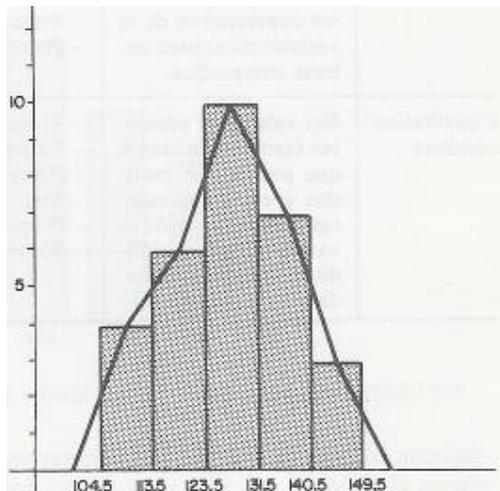


Histogramas.

Consiste en un conjunto de barras o rectángulos dibujados sobre un eje de coordenadas. Cada modalidad se representa por un rectángulo que tiene como base la amplitud del intervalo y como altura la frecuencia de dicho intervalo.

Polígono de frecuencias.

Consiste en una línea quebrada que une los puntos medios de cada intervalo y su frecuencia. Comienza y termina en el eje de abscisas en los puntos medios de los intervalos anterior y posterior a los histogramas.



Actualmente hay programas informáticos, como EXCELL, que permiten realizar de manera muy simple estos gráficos,



Bibliografía

- Carrasco J L. El método estadístico en la investigación médica. 6ª Edición. Editorial Ciencia 3; 1995
- Rodríguez Miñón P. Estadística Aplicada a la Biología. 3ª Edición. Editorial UNED; 1984.
- Polit Dense, Hungler Bernadette. Investigación científica en ciencias de la salud. 6ª edición. Edit McGraw-Hill Interamericana; 2000.
- Fernando Villar et al. Diseño y análisis Epidemiológico. Revista Rol de Enfermería 1987. 112: 13-17.