

Artículo especial

Análisis de los datos en un proyecto de investigación

Begoña Bermejo Fraile

Médico especialista en Medicina Preventiva y Salud Pública. Servicio de Medicina Preventiva y Gestión de la Calidad Hospitalarias. Hospital «Virgen del Camino». Pamplona. Navarra

RESUMEN

Este artículo ofrece una breve descripción de cómo se debe realizar el análisis de los datos de un proyecto de investigación y cómo deben presentarse los resultados. Se hace especial hincapié en los errores más frecuentemente cometidos: gran cantidad de datos irrelevantes, presentación de los mismos resultados en varios formatos (texto, tablas y gráficos), así como el uso exclusivo de la «p» (nivel de significación estadística) como resultado del contraste de hipótesis.

Palabras clave: investigación, metodología, estadística

ABSTRACT

Title: Analysis of the data of a research project.

This article briefly describes the strategy for the analysis of the data collected in a research project and the presentation of the results. The authors especially stress the errors most frequently committed: large amounts of irrelevant data, presentation of the same results in several formats (text, tables, graphs), as well as the utilization of the “p” value (level of statistical significance) alone as the result of hypothesis testing.

Keywords: research, methodology, statistics

Bermejo B. Análisis de datos en un proyecto de investigación. *Matronas Profesión* 2005; 6(3): 30-36.

INTRODUCCIÓN

La sección de resultados de un artículo es la más importante, ya que es donde se expresan, con claridad y

sencillez, los nuevos conocimientos que se están aportando al mundo científico. A menudo, es también la sección más corta, especialmente si el resto de las secciones están bien escritas. En un proyecto de investigación todo tiene que estar planificado, hasta el punto de que el apartado de resultados del artículo puede estar completamente escrito, habiendo dejado sólo los huecos para rellenar con los datos. Esto ayuda al equipo investigador a realizar el análisis en una secuencia lógica, ordenada, utilizando únicamente las pruebas estadísticas necesarias y pertinentes. Por ello, y por sorprendente que parezca, debería ser la parte más sencilla de escribir, ya que se trata de dar respuesta a las preguntas de investigación planteadas. Ninguno de los objetivos planteados se debe quedar sin su correspondiente respuesta en resultados, pero tampoco hay que dar resultados no especificados en los objetivos.

No es preciso proporcionar absolutamente todos los resultados obtenidos y cargar al lector de números. La obsesión por incluirlo todo, sin olvidar nada, no es una prueba de que se dispone de una información ilimitada, sino de que se carece de capacidad de discriminación («El necio colecciona hechos; el sabio los selecciona»).

MATERIAL Y MÉTODOS

Con vistas a la reproducibilidad del estudio, en general es necesario describir los métodos estadísticos con suficiente detalle. Esto se hace en el apartado «Material y métodos», donde se especifica el paquete estadístico utilizado (SPSS, BMDP, SAS, EPIDAT, EPI-INFO), la forma de describir las variables, las pruebas de contraste de hipótesis utilizadas, así como el nivel de significación estadística aceptado, o lo que es preferible, el nivel de confianza de los intervalos.

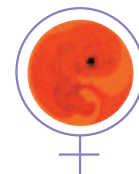
Los paquetes estadísticos, especialmente en determinados análisis multivariados, difieren en la forma de presentar los resultados o incluso en pruebas concretas utilizadas, por ello, es importante especificar cuál se ha utilizado.

La descripción de las variables debe ser coherente con la escala de las variables y con las pruebas de contraste de hipótesis usadas posteriormente. Por ejemplo, no se puede describir una variable con la media y la desviación estándar (entendiendo con ello que sigue una distribución normal) y después emplear pruebas de contraste de hipótesis no paramétricas.

En cuanto al nivel de significación estadística aceptado, también debe especificarse, aunque prácticamente todo el mundo trabaja con un nivel del 5% ($p < 0,05$). De todos modos, es preferible presentar los resultados con el intervalo de confianza, que en este caso sería del 95%.

RECOGIDA DE DATOS

El número y tipo de variables que se recogen en un estudio depende por entero de los objetivos y el diseño



de éste. Hay cierta tendencia a recoger un elevado número de variables, lo que suele traducirse en una menor calidad de los datos y una gran dispersión y confusión en el momento de analizarlos e interpretarlos. La regla es: recoger tantas variables como sean necesarias y tan pocas como sea posible.

Las variables que se recogen pueden agruparse en las siguientes categorías: a) variables que sirven para determinar si un individuo es candidato o no para participar en el estudio. Son los criterios de inclusión y exclusión. Estos datos son importantes para la posterior generalización de los resultados; b) variables universales, que permiten describir a los individuos del estudio, como son la edad, el sexo, estado civil y nivel socioeconómico; c) variables que permiten ver el factor o factores de estudio, y el criterio o criterios de evaluación. Por ejemplo, se quiere ver la influencia de la anestesia epidural en la duración del parto, por lo que se recoge si la mujer ha recibido o no anestesia epidural y la duración del parto; d) variables que pueden actuar como posibles factores de confusión (número de partos previos por ejemplo), modificadores del efecto o pasos intermedios en la cadena causal.

TIPOS DE VARIABLES

Básicamente, existen dos tipos de variables: cuantitativas, que miden una cantidad (peso del recién nacido, edad gestacional, número de embarazos) y cualitativas o categóricas, que miden una cualidad (como el sexo, raza, presencia o no de una enfermedad). Hay un tercer tipo de variables, las semicuantitativas u ordinales, que reciben su nombre por estar a caballo entre las cuantitativas y las cualitativas. Si, por ejemplo, se mide el dolor en grados 1, 2, 3 y 4, de tal modo que el nivel del mismo es cada vez mayor, no se puede afirmar que quien tiene un dolor grado 4 experimenta exactamente el doble de dolor que quien expresa que tiene un grado de dolor 2 o cuatro veces más que quien refiere un dolor grado 1. Sólo se puede afirmar que el grado 4 es mayor que el 2, pero no cuánto mayor. Se denomina categorizar una variable cuantitativa a convertirla en cualitativa u ordinal. Por ejemplo, se puede recoger el número de cigarrillos que una persona fuma diariamente, o categorizarla en no fumador, fumador de 1 a 10 cigarrillos, de 11 a 20 cigarrillos y más de 20, o incluso categorizarla en sólo dos categorías, fumador y no fumador. En términos generales, lo más conveniente es recoger las variables en toda su precisión (cuantitativa), porque siempre estamos a tiempo de categorizarla. Recogerla desde el principio en intervalos es perder mucha información que, posteriormente, podría ser muy valiosa.

Para introducir las variables cualitativas en la base de datos, se recomienda registrarlas de forma codificada,

por ejemplo: 1: sí / 2: no, evitando el dígito cero para no confundirlo con los datos ausentes. De hecho, es conveniente prever un código para estos datos desconocidos. Puede ser 9, 99, 999, siempre que estos valores no puedan coincidir con un valor real.

DESCRIPCIÓN DE LAS VARIABLES

Antes de comenzar con el análisis estadístico, es muy importante depurar la base de datos: intentar cumplimentar algunos datos ausentes y corregir los datos incongruentes (una hemoglobina de 35 mg/dL o una fecha del alta previa a la del ingreso).

Una vez hecho esto, se comienza con la descripción de los datos de la muestra, pero de la muestra real a la que se refieren los resultados. Generalmente, no todas las variables están recogidas en todos los individuos y esto es especialmente problemático si estas pérdidas están relacionadas con las variables de interés en el estudio. De ahí la importancia no sólo de cuantificar las pérdidas, sino también de intentar averiguar sus causas e indicarlas. La descripción de la información recogida sobre una variable cualitativa se realiza mediante la distribución de frecuencias de cada una de las categorías, indicando tanto la frecuencia absoluta como el porcentaje. Para las variables dicotómicas (cualitativas de dos categorías), normalmente es suficiente con indicar la frecuencia de una de las categorías, ya que se sobrentiende que el resto corresponde a la categoría restante. Por ejemplo, al describir el sexo de las personas que integran una muestra de 60 mujeres y 140 hombres, es suficiente decir que el 30% (60) son mujeres. La forma ideal de presentación gráfica de este tipo de variables es mediante el diagrama de sectores (pastel) o el de barras, donde cada sector o barra representa la frecuencia relativa de dicha categoría en el total de la muestra.

En algunos casos, la descripción de una variable cuantitativa también se puede hacer mediante la distribución de frecuencias. Se trata de realizar intervalos (por ejemplo, dividir la edad en grupos de 10 años) y ofrecer la frecuencia de cada uno de los intervalos. La descripción de una variable cuantitativa es más correcta mediante las medidas de tendencia central y de dispersión. Las medidas de tendencia central informan del valor de la variable en torno al cual se agrupan los restantes valores. Son tres: la media, la mediana y la moda. Para el cálculo de la media aritmética, se suman todos los valores observados y se divide por el total de observaciones. La mediana es el valor que divide a la muestra en dos partes iguales, de modo que la mitad de los individuos tienen valores menores que la mediana y la otra mitad mayores. Y, por último, la moda, que es el valor más frecuente, el que más veces aparece. Si en una muestra de estudio, la media de la edad es de 30 años, y todos los individuos tienen entre 28 y 33 años, y en otra muestra, la media también es de 30

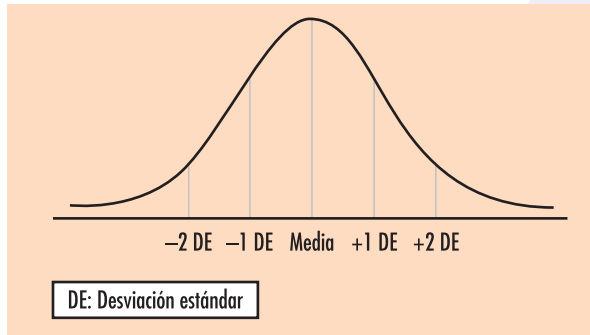


Figura 1. Campana de Gauss

años, pero las edades están comprendidas entre 20 y 40 años, aunque las dos muestras presentan la misma media, la variabilidad de los datos es diferente.

Por tanto, a la hora de describir una variable, no es suficiente con dar una medida de tendencia central, sino que también hay que informar de la dispersión de estos datos, de si están todos muy agrupados o no alrededor de la media. Éstas son las medidas de dispersión, entre las que destacan la desviación estándar y la amplitud o rango. El rango es la diferencia entre el valor máximo y mínimo en la muestra, y la desviación estándar, sin entrar en los detalles de su cálculo, también nos informa de la variabilidad de la variable (valga la redundancia), de modo que, en una variable que sigue una distribución simétrica (distribución normal, la famosa campana de Gauss), el 95% de los valores se encuentra entre la media más y menos 1,96 desviaciones estándar (figura 1). En lugar de 1,96 se suele redondear a 2.

En una muestra de 100 recién nacidos, el peso medio es de 3.200 g y la desviación estándar (DE) 150 g. Si el peso sigue una distribución normal, sólo con estos datos se sabe que el 95% de los recién nacidos están entre 2.900 y 3.500 g ($3.200 \pm 2 \cdot 150$). Si la desviación estándar fuese mayor, por ejemplo 400 g, quiere decir que el peso puede variar en un rango mayor de valores. En este último caso, el 95% de los recién nacidos tendría un peso entre 2.400 y 4.000 g.

En términos generales, si una variable sigue una distribución normal, se describe con la media y la desviación estándar, y si no sigue tal distribución, se describe con la mediana, valor mínimo y máximo. Hay variables que pueden parecer cuantitativas y, sin embargo, no lo son. Por ejemplo, la medición del dolor en una escala de 0 a 100. Esta variable, aparentemente, es cuantitativa, sin embargo es ordinal, porque si un paciente refiere un nivel de dolor 20, no quiere decir que tenga exactamente el doble de dolor que quien refiere un nivel 10, sólo que tiene más dolor. No mide una cantidad real, sino que establece un orden, un rango de valores. No es lo mismo pasar de 0 a 10, que de 50 a 60 por ejemplo. Estrictamente, este tipo de variables, al no ser cuantitativas, se deberían describir con la mediana, valor mínimo y máximo.

ESTIMACIÓN DE PARÁMETROS

Para conocer la realidad (objetivo de la investigación), lo ideal sería poder estudiar poblaciones completas, pero a veces son tan grandes que debemos limitarnos al estudio de una porción de éstas, las muestras. Esto es lo que hace la estadística inferencial: sacar conclusiones de un gran número de acontecimientos a partir del estudio de una parte de ellos. Un ejemplo son los sondeos de opinión pública, en los que se entrevista a una pequeña proporción de la población para un objetivo concreto (por ejemplo, qué votarán en las próximas elecciones). Por el simple hecho de estudiar muestras y no poblaciones, por puro azar, siempre existe el riesgo de cometer un error en la estimación (de ahí que este error se denomine error aleatorio).

Siguiendo con el ejemplo previo, se quiere estimar el peso medio de los recién nacidos y para ello se estudia una muestra de 200 recién nacidos; en esta muestra, el peso medio ha sido de 3.200 g y la DE de 150 g. Se podría haber seleccionado cualquier otra muestra y que el peso medio hubiese sido 3.000 g u otra donde el peso medio fuese de 3.500 g. De hecho, se podría haber estudiado cualquiera de las miles de posibles muestras de 200 recién nacidos, cada una con su media. ¿Cuál es la media real? De no estudiar absolutamente a todos los recién nacidos, no es posible saberlo. Por ello, lo que se hace es estimar un rango de valores entre los que confiamos que se encuentre la verdadera media. Este rango de valores es lo que se denomina el intervalo de confianza (IC). Este intervalo de confianza del 95% (el más frecuentemente utilizado) se calcula del siguiente modo:

$$IC\ 95\% = \text{estimado} \pm 1,96\ EE$$

Donde EE es el error estándar y cuya fórmula varía en función del tipo de parámetro estimado:

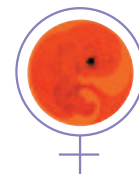
El EE de una media (cuando la variable en estudio es cuantitativa, como el peso de los recién nacidos) es:

$$EEM = \frac{DE}{\sqrt{n}}$$

El EE de una proporción (cuando la variable en estudio es cualitativa, como fumar o no fumar) tiene la siguiente fórmula:

$$EEP = \sqrt{\frac{p \cdot q}{n}}$$

donde p es la proporción en estudio, en este caso la de fumadores, q la proporción complementaria, la de no fumadores (1-p) y n el tamaño de la muestra.



Ejemplo: se han estudiado 400 mujeres en edad fértil y el nivel medio de hemoglobina es 12,2 mg/dL y la DE 2,8 mg/dL. ¿Cuál será la media poblacional?

$$IC\ 95\% = 12,20 \pm 1,96 \frac{2,8}{\sqrt{400}} = 12,20 \pm 0,2744\ (11,93 - 12,47)$$

Confiamos (por ello se llama intervalo de confianza) en que la verdadera media de hemoglobina se encuentre en ese rango de valores, entre 11,93 y 12,47 mg/dL.

Si en lugar de estudiar 400 mujeres, se hubiesen incluido sólo 20, el intervalo de confianza de la estimación sería:

$$IC\ 95\% = 12,20 \pm 1,96 \frac{2,8}{\sqrt{20}} = 12,20 \pm 1,2272\ (10,97 - 13,43)$$

En este caso, el intervalo es mayor, porque, a menor tamaño de la muestra, mayor es el error en la estimación, y a mayor tamaño menor error, hasta el punto de que si se estudiara la población completa el error sería nulo.

Un ejemplo de estimación de una proporción sería el caso en el que se estudia una muestra de 200 mujeres entre 30 y 40 años para conocer la prevalencia de fumadoras. Ha resultado un 25%.

$$IC\ 95\% = 12,20 \pm 1,96 \sqrt{\frac{0,25 \cdot 0,75}{200}} = 0,25 \pm 0,03\ (0,19 - 0,31)$$

Con un 95% de probabilidades, la verdadera prevalencia de fumadoras entre las mujeres de 30 a 40 años está entre el 19 y el 31%.

Las estimaciones puntuales sirven de muy poco, siempre tienen que ir acompañadas de su intervalo de confianza. Si es un intervalo estrecho, porque el tamaño de la muestra es importante, el resultado estará aportando información. Pero si el tamaño de la muestra es pequeño, se han estudiado muy pocos pacientes, el intervalo será muy amplio y no aporta información relevante. Imagínese que la prevalencia de ferropenia en las mujeres en edad fértil es del 40% (IC 95%: 5 a 75%). Decir que la prevalencia está entre el 5 y el 75% es como no decir nada. A esta misma conclusión se llegaría sin necesidad de realizar un estudio. Es preciso distinguir la DE del EE. La desviación estándar es un número que refleja la variabilidad de los individuos (no todos los recién nacidos tienen el mismo peso ni todas las mujeres la misma hemoglobina). Si la variable sigue una distribución normal, el 95% de los individuos se encuentran entre la media \pm 1,96 DE. El EE es un número también, pero que refleja el error del muestreo, el error debido al azar en la selección de la

muestra. Sirve para el cálculo del intervalo de confianza mediante las fórmulas vistas con anterioridad.

Es un error relativamente frecuente describir los datos con la media y el EE, en lugar de con la DE. Suele ocurrir cuando la distribución de la variable es muy sesgada y la DE es muy grande; entonces se presenta el EE, que siempre es un número más pequeño (porque se divide por la raíz cuadrada del tamaño muestral). En última instancia, la presentación de los resultados depende de los objetivos del estudio: si se trata de describir una variable cuantitativa, se presentan la media y la DE; y si se trata de estimar un parámetro en la población, lo conveniente es determinar el intervalo de confianza de dicha estimación, para cuyo cálculo se utiliza el EE.

CONTRASTE DE HIPÓTESIS

La estadística inferencial, además de la estimación de parámetros, se encarga del contraste de hipótesis. Si se plantean preguntas del tipo: ¿Es más alto el porcentaje de cesáreas en los hospitales públicos que en los privados? ¿La vigilancia electrónica del feto durante el trabajo de parto, tiene efectos positivos en el resultado del embarazo? ¿La lactancia materna disminuye el número de infecciones en los lactantes?

El contraste de hipótesis consiste en comprobar si una hipótesis puede ser aceptada o rechazada con cierta probabilidad de acertar. La hipótesis que va a ser contrastada es la hipótesis nula, la hipótesis de la igualdad, de la ausencia de diferencia. Se formula con la intención expresa de ser rechazada. Si se rechaza, puede entonces aceptarse la hipótesis alternativa, que es la hipótesis de investigación del experimentador. Por ejemplo, el objetivo de nuestro estudio es ver si el peso de los recién nacidos es menor entre las madres fumadoras respecto a las no fumadoras. La hipótesis nula es que no hay diferencias entre el peso de los recién nacidos según la madre sea fumadora o no, y la hipótesis alternativa es que el peso de los recién nacidos de madres fumadoras es menor que el peso de los recién nacidos de madres no fumadoras. Siempre existe una posibilidad de error. Si se rechaza la hipótesis nula siendo cierta, el error se denomina alfa, o error tipo I, y si no se rechaza la hipótesis nula siendo falsa, el error se denomina beta o error tipo II. Se puede plantear el contraste de hipótesis como un juicio: se trata de saber si un individuo es culpable de asesinato; ante la ausencia de pruebas, la premisa es la presunción de inocencia (hipótesis nula). Si se encuentra un número suficiente de pruebas (diferencias significativas), se declara al individuo culpable (hipótesis alternativa). Si no se encuentran estas pruebas, se declara al individuo inocente. Se llegue a la conclusión que se llegue, siempre existe la probabilidad de equivocación; se trata de que esta probabilidad de error sea muy pequeña.

A continuación, sin entrar en detalle, se especifican las pruebas de contraste de hipótesis utilizadas en función de la asociación que se pretende estudiar y el tipo de variables que entran en juego (tabla 1).

Comparación de proporciones

Ejemplo: ¿es mayor el porcentaje de episiotomías entre las mujeres que paren en casa que entre aquellas que lo hacen en el hospital? Comparar proporciones es lo mismo que estudiar la asociación entre dos variables cualitativas. En este caso las variables son: episiotomía (sí o no) y lugar del parto (casa u hospital). La prueba de contraste de hipótesis utilizada es la ji al cuadrado.

Comparación de medias

Ejemplo: se ha estudiado la presión sistólica en un grupo de mujeres de 35 a 40 años, que toman anticonceptivos orales, y ha resultado una presión media de 133 mmHg y una DE de 15 mmHg. Se ha comparado con un grupo de mujeres de la misma edad que no toman anticonceptivos orales. En éstas la presión ha sido de 127 mmHg y la DE de 18 mmHg. La pregunta puede plantearse de los siguientes modos: ¿es diferente la presión sistólica en las mujeres que toman anticonceptivos orales respecto a las que no los toman?, ¿aumentan los anticonceptivos orales los niveles de presión sistólica?, ¿están asociados los anticonceptivos a la presión sistólica?

Se están comparando dos medias, la media de presión en las mujeres que toman anticonceptivos con la media de presión en las que no toman. Hay, por tanto, dos variables: una cualitativa de dos categorías (tomar o no anticonceptivos orales) y una cuantitativa (la presión arterial sistólica). La hipótesis nula es que ambas medias son iguales y la hipótesis alternativa es que ambas medias son diferentes. La comparación de dos medias se realiza mediante la prueba *t* de Student.

El análisis de la variancia (ANOVA) se utiliza para la comparación de más de dos medias. Tenemos nuevamente dos variables: una cualitativa (que define los grupos y, en este caso, tiene más de dos categorías) y una cuantitativa cuyas medias pretendemos comparar. Por ejemplo, se comparan los días de estancia hospitalaria (variable cuantitativa) según el parto haya sido vía vaginal no instrumentado, vía vaginal instrumentado o por la realización de cesárea (variable cualitativa de tres categorías).

Correlación

En estadística, se entiende por correlación el estudio de la asociación entre dos variables cuantitativas.

Ejemplos de este tipo de análisis son: ¿aumenta la presión arterial a medida que se incrementa la cifra de cigarrillos fumados?, ¿existe alguna correlación entre la edad y el colesterol? El estadístico que se utiliza es el coeficiente de correlación *r* de Pearson. Su intervalo de valores va de -1 a $+1$, siendo el cero el valor nulo.

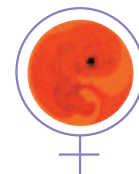
Cuanto más cercano a $+1$, mejor correlación, y positiva (al aumentar una variable aumenta la otra, al aumentar el número de cigarrillos fumados aumenta la presión arterial), y cuanto más cercano a -1 , mejor correlación también, pero negativa (al aumentar una variable disminuye la otra así, por ejemplo, al aumentar la duración de la lactancia materna disminuyen los niveles de ferritina sérica). Un coeficiente de cero indica que no existe correlación entre las variables.

Análisis de datos apareados

Por poner un ejemplo muy sencillo, se puede decir que datos apareados son aquellos en que la misma variable es medida en la misma persona antes y después de una intervención.

Tabla 1. Pruebas estadísticas más utilizadas según tipo de variables

	<i>Cualitativa, dos categorías</i>	<i>Cualitativa, más de dos categorías</i>	<i>Ordinal</i>	<i>Cuantitativa</i>
Cualitativa, dos categorías Datos independientes	Ji al cuadrado	Ji al cuadrado	U de Mann-Whitney	<i>t</i> de Student
Cualitativa, dos categorías Datos emparejados	McNemar	Q de Cochran	Wilcoxon	<i>t</i> de Student para datos apareados
Cualitativa, más de dos categorías Datos independientes	Ji al cuadrado	Ji al cuadrado	Kruskal-Wallis	Análisis de la variancia
Cualitativa, más de dos categorías Datos apareados	Q de Cochran	Q de Cochran	Friedman	ANOVA para medidas repetidas
Cuantitativa	<i>t</i> de Student	Análisis de la variancia (ANOVA)	Correlación Spearman	Correlación de Pearson



Comparación de proporciones para datos apareados

Ejemplo: se pretende evaluar el cambio en la percepción de los pacientes sobre la atención sanitaria, después de haber estado ingresados en un hospital durante 10 días. Se seleccionan aleatoriamente 100 pacientes y se les pregunta qué opinan de la atención sanitaria el primer día del ingreso y en el momento del alta. Se pretende valorar si ha habido un cambio de opinión durante el ingreso, si la proporción de los que ha mejorado su opinión es diferente de la proporción de los que ha empeorado su opinión. Para ello, se realiza la prueba de McNemar. Cuando una de las variables cualitativas tiene más de dos categorías se utiliza la Q de Cochran.

Comparación de medias para datos apareados

Cuando en una población se selecciona una muestra de individuos y se realizan dos mediciones de una misma variable (por ejemplo, hemoglobina antes y después de un tratamiento con hierro), para comprobar si se ha producido un cambio significativo se realiza la prueba *t* de Student para datos apareados. Se calcula la variable «diferencia» de mediciones, restando, para cada individuo, el valor previo del posterior (o al contrario, ya que lo importante es la diferencia) y se comprueba si esta media es diferente (significativamente diferente) de cero. Cuando son más de dos las mediciones de la variable cuantitativa, en varios periodos, el análisis que se realiza es ANOVA para medidas repetidas.

Pruebas no paramétricas

Hasta ahora, se han explicado lo que se denominan pruebas paramétricas, cuando las variables cuantitativas siguen una distribución conocida (fundamentalmente la normal). Cuando, por el contrario, las variables no siguen una distribución normal (están muy sesgadas, son asimétricas) o son variables ordinales, se deben utilizar los métodos no paramétricos.

Las pruebas no paramétricas más comunes son: la prueba U de Mann-Whitney, equivalente no paramétrico a la *t* de Student para muestras independientes; la prueba de Kruskal-Wallis, equivalente no paramétrico al ANOVA (análisis de la variancia); el equivalente no paramétrico al ANOVA para medidas repetidas es el test de Friedman; la prueba de rangos con signo (prueba de Wilcoxon), equivalente no paramétrico a la *t* de Student para datos emparejados; el coeficiente de correlación *r* de Spearman, equivalente no paramétrico al coeficiente de correlación *r* de Pearson.

ANÁLISIS MULTIVARIABLE

En un estudio, en el que se comparan los resultados del parto según éste haya sido atendido por matronas o médicos, se observa que la tasa de complicaciones en el parto es menor en aquellos atendidos por matronas. Pero, también es conocido que, en términos generales, las matronas atienden más los partos de bajo y moderado riesgo obstétrico. Por tanto, ¿a qué se deben esos

mejores resultados, a que atienden mejor el parto o a que ya de entrada los partos son de menor riesgo? Aquí, la relación entre las distintas variables que influyen en el resultado del parto nos está confundiendo. Con el análisis multivariable se pretende obtener el efecto de cada una de las variables, ajustando por los factores de confusión. Destacan la regresión lineal múltiple y la regresión logística, según la variable dependiente (la variable resultado) sea cuantitativa o dicotómica, respectivamente.

RESULTADOS DEL CONTRASTE DE HIPÓTESIS

Muy frecuentemente, el resultado del contraste de hipótesis se expresa mediante la *p* (nivel de significación estadística o probabilidad asociada al error tipo I), de modo que si la *p* es $<0,05$, decimos que las diferencias son estadísticamente significativas, mientras que si es mayor, las diferencias encontradas no son estadísticamente significativas.

Por un lado, no existe una lógica detrás del hecho de que prácticamente todo el mundo utilice un nivel de significación estadística del 5%. Por otro lado, el tamaño de la muestra, en estadística, es como una lupa: cuando el tamaño de las muestras es grande, cualquier pequeña diferencia se detecta como estadísticamente significativa y, en cambio, cuando el tamaño de las muestras es pequeño, grandes diferencias pueden no resultar estadísticamente significativas. De ahí la importancia del tamaño muestral; pero más importante es todavía recalcar que una asociación estadísticamente significativa no quiere decir que sea clínicamente relevante. La «*p*» no proporciona en absoluto información de la magnitud del efecto. Una $p=0,001$ no quiere decir que el efecto es mayor que si ha resultado $p=0,78$. Mediante un ejemplo se ilustra mejor este tema. Se comprueba en un estudio pequeño que un medicamento disminuye la presión arterial diastólica en media 30 mmHg, no resultando estas diferencias estadísticamente significativas ($p=0,32$). En cambio, en otro estudio con un número de pacientes muchísimo mayor, se observa que otro medicamento disminuye la presión arterial diastólica en media 6 mmHg, siendo estas diferencias estadísticamente significativas ($p=0,01$). Así, a simple vista, ¿cuál le parece mejor medicamento hipotensor?

De aquí se deriva también que, en las pruebas de contraste de hipótesis, no es suficiente con presentar la *p*, porque no ofrece una idea de las diferencias observadas. Lo más conveniente es presentar los intervalos de confianza, el rango de valores probable, para que el lector, con un espíritu crítico, pueda interpretar directamente esos resultados. Además, si el intervalo de confianza de las diferencias observadas incluye el valor nulo, quiere decir que las diferencias no son estadísticamente significativas, y si el intervalo de confianza no incluye el valor nulo, las diferencias observadas son estadísticamente significativas.

Si se comparan proporciones, se presenta el intervalo de confianza de la diferencia de proporciones; si se comparan medias, el intervalo de confianza de la diferencia de las medias; hay otras medidas específicas como el riesgo relativo (RR) en los estudios de incidencia, la *odds ratio* (OR) (en castellano «tasa de probabilidad») en los de prevalencia, o la reducción relativa del riesgo (RRR), la reducción absoluta del riesgo (RAR) o el número que es necesario tratar (NNT) en los ensayos clínicos; en todos los casos, estas medidas deben ir acompañadas de su intervalo de confianza.

ELABORACIÓN DE TABLAS Y GRÁFICOS

En general, no hay que presentar los mismos datos en más de una forma, por escrito, en tablas y en gráficos. Algunos autores creen que una tabla o una figura añade importancia a los datos, y no es así; no se trata de adornar los datos, se trata de elegir la manera más clara y evocadora de presentar la información. Es importante recalcar esto porque, además, el coste de publicar tablas y gráficos es muy elevado en comparación con el del texto, por lo que el número de cuadros o figuras que se pueden enviar viene frecuentemente limitado desde la revista.

Cuando la lista de datos es importante, y en el texto resultaría muy enmarañado, se elige una tabla, omitiendo, por ejemplo, los datos de variables secundarias que no son relevantes, o datos sobre variables con muy pocos casos en alguna de sus categorías.

La tabla debe estar muy bien ordenada y estructurada, y es preferible que los datos se lean de arriba abajo que transversalmente. Debe revisarse muy bien el contenido, de modo que los números cuadren, y, cuando sea pertinente, sumen el 100%. También es importante comprobar que no hay contradicciones con datos de las tablas que pueden aparecer en el texto, en el resumen o en un gráfico.

El título, los encabezamientos, deben ser lo suficientemente claros para que el significado de los datos sea comprensible sin necesidad de acudir al texto. Cualquier abreviatura o sigla empleada en la tabla debe llevar su significado al pie de ella, a pesar de haberse detallado en el texto.

Si se trata de elegir entre gráfico y tabla, la decisión puede depender de si se quieren presentar datos numéricos exactos, en cuyo caso es preferible una tabla, o simplemente tendencias y distribución de los datos, que se verán mejor en un gráfico.

ESTILO

Si hay alguna cualidad importante de la sección de resultados es la claridad; claridad como sinónimo de inteligibilidad. El autor no sólo tiene que entenderse a sí mismo, sino que tiene que hacerse entender por los demás. También es importante ser conciso, se trata de decir todo lo que hay que decir sin utilizar datos o palabras innecesarias (lo breve si bueno, dos veces

bueno). La sección de resultados varía entre 1 y 4 páginas, y sólo deben exponerse los datos, no se deben comentar o discutir, ya que para ello hay otras secciones. No hay que utilizar un lenguaje rebuscado, frases complicadas, vacías o palabras de relleno. Es preferible decir «de ahí que» o «dado que» en lugar de «de lo anteriormente expresado se deduce que» o «habida cuenta del hecho de que». Resulta mucho más atractivo leer a un autor que, además de decir cosas interesantes, va directamente al grano.

Pero tampoco hay que irse al otro extremo y convertir el artículo en un laberinto de abreviaturas. Se debe decir el nombre completo la primera vez que se utiliza una abreviatura, por ejemplo DE (desviación estándar), aunque esta abreviatura, junto con la del error estándar (EE), están tan extendidas que podría obviarse el nombre completo incluso la primera vez. Aprovecho la ocasión para advertir del incorrecto, pero muy frecuente uso del símbolo \pm para introducir la DE. Por ejemplo, se han estudiado 200 recién nacidos, cuyo peso medio al nacer fue de 3.150 g (\pm 210). La forma correcta sería «...cuyo peso medio fue de 3.150 g (DE= 210)». Este error proviene del hecho de que cuando una variable sigue una distribución normal, el 95% de los valores se encuentra entre la media más y menos 1,96 desviaciones estándar. Otra característica es la precisión del lenguaje; en lo que se refiere a los términos estadísticos, hay que tener mucho cuidado con el empleo no técnico de términos como «azar», «correlación», «significativo», «normal» o «muestra». Por último, sobra decir que una de las características fundamentales del buen investigador es la honradez, lo que redundará en su credibilidad. Hay que exponer todos los resultados relevantes de la forma más objetiva posible, tanto si van a favor como en contra del punto de vista del autor.

BIBLIOGRAFÍA

- Armitage P, Berry G. Estadística para la investigación biomédica. Barcelona: Doyma, 1992.
- Day RA. Cómo escribir y publicar trabajos científicos, 2.^a ed. Washington: Organización Panamericana de la Salud, 1996.
- Icart MT, Torrents ML, Bermejo B, Canela J. Enfermería Comunitaria. Epidemiología. Barcelona: Masson, 1997.
- Rozman C, For M, Herranz G, Gracia D, Lience E, Pulido M, et al. Manual de estilo. Barcelona: Doyma, 1995.
- Rosner B. Fundamentals of biostatistics. Belmont: Duxbury Press, 1995.
- Siegel S. Estadística no paramétrica. México: Trillas, 1991.
- Silva LC. Cultura estadística e investigación científica en el campo de la salud: una mirada crítica. Madrid: Díaz de Santos, 1997.

Correspondencia

Begoña Bermejo Fraile
Servicio de Medicina Preventiva y Gestión de la Calidad Hospitalarias
Hospital «Virgen del Camino». Pamplona. Navarra
begona.bermejo.fraile@cfnavarra.es