Revista Mexicana de Cardiología

Volumen 25, Número 2
Abril - Junio 2014
pp 129 - 131

Comunicado breve

Estadística descriptiva y selección de la prueba

Cuauhtémoc Acoltzin Vidal*

Las finalidades de la estadística son: sintetizar los datos, estimar y hacer inferencia a la población de referencia y ajustar los datos según la influencia de factores de confusión.

En todos los métodos se supone que la muestra es un subgrupo estadístico de la población de la que se ha extraído, esto quiere decir que todas las mediciones de la población tienen la misma probabilidad de estar incluidas en la muestra; por lo tanto, el azar dicta cuáles de las mediciones se incluyen realmente. La selección aleatoria de las mediciones incluidas en la muestra determina cuánto se aproxima el estadístico, es decir, el valor numérico obtenido de la muestra, al valor real del parámetro poblacional.

La síntesis permite expresar con dos o tres cifras la distribución de los datos, tanto en su tendencia a agruparse en un punto central, como en su dispersión.

1. Si la distribución es normal (lo que ocurre sólo en población general) o se parece a la normal (en forma de campana) se expresan con el valor medio (o media aritmética) y la desviación estándar.
2. Si la distribución no es normal, pero la naturaleza de los datos es numérica, la expresión incluye la mediana y la moda, que indican la agrupación en el centro y la predominante y el recorrido intercuartílico que expresa dispersión (ubicación de los datos correspondientes al inicial, a los correspondientes al 25% y al 75% del número de observaciones; y se complementa con la mediana, que es el 50%).

Para hacer estimación, en especial cuando los datos son nominales (es decir, que se nombran, pero no se pueden medir porque sólo se sabe su presencia o su ausencia y, por lo tanto, se cuentan) se recurre a:

1. El rango (que es la amplitud entre la observación más grande y la menor se expresa como un número. No debe confundirse con el intervalo o la clase que son los límites inferior y superior).
2. Al cálculo de razón (número de observaciones con un resultado [a] entre el número de observaciones con otro [b] con la fórmula a/b). Proporción a/(a + b) o porcentaje (proporción multiplicada por 100).
3. A los índices o tazas:

    a. De prevalencia (cuya fórmula es semejante, pero en el denominador se usa un multiplicador [base] y un espacio de tiempo específicos que por lo general son 10,000 y un año. Éste se expresa como tiempo-persona)
    b. O de incidencia (que se calcula en una cohorte dinámica a lo largo de un periodo y, por ello, se deben tener en cuenta los cambios que ocurren en numerador y denominador –debidos a ingreso y pérdida de sujetos en éste–).

4. Ubicación por percentiles –o porcentiles–, es decir, valores porcentuales (o de peso) de cada dato según el número de veces que se presenta entre todas las observaciones (se calculan dividiendo el número de observaciones por dato entre «n» [número de observaciones]). Es acumulativo, es decir, se va sumando a los valores previos hasta completar 1, que es 100%).

La inferencia estadística –o prueba de significancia estadística– permite saber qué tanto se espera que varíe el estadístico en relación con el valor hipotético del parámetro poblacional sobre la base de la variabilidad del azar entre las muestras aleatorias. El valor hipotético con el que se comparan las estimaciones se llama hipótesis nula. El objetivo de las pruebas de significación estadística es calcular el valor p, que es la probabilidad de que la hipótesis nula

* Médico cirujano, cardiólogo y maestro en ciencias médicas. Universidad de Colima. Profesor de Estadística aplicada a la investigación en medicina. Maestría en Ciencias Médicas, Facultad de Medicina, Universidad de Colima.

Rev Mex Cardiol 2014; 25 (2): 129-131

sea cierta, es decir, que dos subconjuntos del conjunto universal estén unidos, teniendo una muestra que sea como mínimo tan distinta de la indicada por la hipótesis nula como la realmente obtenida si aquélla –la hipótesis nula– realmente describe a la población.

Para extrapolar los resultados a la población universal se recurre al cálculo del intervalo de confianza o estimación por intervalo. Lo que se refiere a la posible variación de un valor, por aumento o por disminución, es decir, en sentido bilateral (aunque hay ocasiones en que sólo se espera variación en un sentido, es decir, es unilateral) al hacer nuevas mediciones de la población experimental. Por lo general, se calcula para extrapolar al 95% de la población universal ya que es la representada por 1.96 desviaciones estándar a partir de la media (llamada valor Z).

El ajuste de los datos es necesario cuando hay una variable dependiente (VD) y dos o más variables independientes (VI) que se pueden medir en la misma escala o en diferentes; éste permite:

1. Investigar la relación que hay entre una VD y una VI mientras controla el efecto de otras VI.
2. Realizar pruebas de significación estadística de diversas variables, manteniendo al mismo tiempo la probabilidad (alfa) de cometer un error tipo I (rechazar falsamente la hipótesis nula).
3. Comparar por separado la capacidad de dos o más VI para estimar los valores de una VD.

Antes de seleccionar un método estadístico se deben tomar dos decisiones:

1. ¿Cuál es la variable dependiente y cuál la variable independiente?
2. ¿Qué tipo de datos constituyen cada una de esas variables?

La VD puede identificarse como la de interés analítico (o el desenlace principal del estudio). Identificarla es el primer paso para seleccionar la prueba estadística. La VI es aquélla cuyo comportamiento es conocido a priori o se puede controlar a discreción.

Puede suceder que no haya ninguna VI o que se incluya una o más. El número de variables independientes determina el tipo de análisis estadístico que es apropiado para analizar los datos.

Se califican como univariante si no hay VI, bivariante si hay una VD y una VI, multivariante si hay una VD y varias VI.

Además existe el diseño multivariado con varias variables dependientes y varias variables independientes; si se desea considerarlas simultáneamente en el mismo estudio se enfrentarán dificultades en el método y en la interpretación que hacen conveniente la participación de un experto en matemáticas y en estadísticas, que entienda y congenie con el investigador. Se ha vuelto costumbre unir varias VD en una sola llamada desenlace final primario y después analizar cada una por separado como desenlaces finales secundarios o terciarios, lo que simplifica el análisis porque se vuelven nominales.

Los datos son los resultados de las mediciones de los sujetos de estudio en la muestra. Para seleccionar la prueba estadística se debe definir la distribución y el tipo de datos que constituyen las mediciones de cada variable.

Para definir la distribución se ordenan los resultados de menor a mayor, y se hace una curva para distribuir la frecuencia con que se presentan. Si ésta es curva normal, paramétrica o parecida a la normal se aplicarán las pruebas paramétricas. Si la curva no es simétrica se dice que su distribución es libre, no paramétrica o con sesgo y se usarán las llamadas pruebas no paramétricas.

La altura de la curva puede variar. Esto es útil en el sentido de mostrar que la diferencia mínima de un caso puede cambiar la absoluta normalidad de una curva: una sola medición puede hacer que la curva normal y la curva obtenida difieran por lo que, para aplicar las pruebas de significancia, se utiliza un criterio llamado grados de libertad, el cual consiste en quitar uno al número de datos obtenidos, concediendo así la posibilidad de variación mínima en comparación con la curva normal.

Con el fin de seleccionar la técnica estadística e interpretar el resultado se distingue entre tres categorías de datos (según el nivel de medición de las variables de que proceden):

a) Continuos
b) Ordinales
c) Nominales

Los datos continuos se pueden observar entre un número infinito de valores espaciados regularmente entre dos puntos cualquiera de su intervalo de medidas. Ocupan cualquier lugar en la recta numérica. Se caracterizan porque se pueden medir. Los ordinales no se pueden fraccionar porque siguen un orden fijo (1º, 2º, 3º, etc. por ejemplo, y se hallan en mediciones antes y después o pareadas); y en los nominales sólo se identifica su presencia o ausencia y se tienen que contar.

Rev Mex Cardiol 2014; 25 (2): 129-131

Por fin se decide si entre la VD nominal y la VI nominal se buscará asociación con las pruebas χ^2. Si La VD es nominal, la VI continua y la distribución es normal se buscará diferencia de medias con pruebas paramétricas; si la distribución es libre se recurrirá a pruebas no paramétricas que comparan distribuciones entre los subgrupos observados. Si ambas variables son continuas se buscará relación con pruebas de correlación y determinación.

BIBLIOGRAFÍA COMPLEMENTARIA

1. Dawson-Saunders B, Trapp RG. Bioestadística médica. México, D.F: Editorial El Manual Moderno S.A. de C.V.; 1990: 67.
2. Colimón KM. Fundamentos de epidemiología. Madrid: Edit Díaz de Santos S.A; 1990.
3. Riegelman RK, Hirsh RP. Cómo estudiar un estudio y probar una prueba: Lectura crítica de la literatura médica. 2a ed. Washington, D.C.: O.P.S.; 1992: 173-234.

Dirección para correspondencia:

Cuauhtémoc Acoltzin Vidal
Calzada del Campesino Núm. 99,
28060, Colima, Colima.
Teléfono y fax: (01312) 3 13 66 09
E-mail: cuauhtemoc_acoltzin@ucol.mx