

Conceptos básicos de estadística descriptiva útiles para el médico

Pilar Mata Miranda,¹ Francisco Reynoso Arizmendi,² Alberto Salazar Valadez³

RESUMEN

Una cantidad importante de investigaciones médicas requieren de estadística descriptiva, cuya utilidad radica, como su nombre lo dice, en describir las características de la población que estudiamos. La descripción de la población gira en función de la distribución de las variables. De acuerdo con lo anterior, para la descripción de una población cuyas variables sean en escala de intervalo o de razón, pueden usarse las medidas de tendencia central y las medidas de dispersión. Algunas de las medidas de frecuencia de la enfermedad más utilizadas son la razón, la proporción y la tasa.

Palabras clave: Estadística, estadística descriptiva, media, mediana, desviación estándar, razón, proporción, tasa.

INTRODUCCIÓN

Como profesionales de la salud, dedicados a la asistencia médica, es difícil mantener un estado de actualización que acerque a métodos que tradicionalmente se han dejado a profesionistas de otras áreas como de las ciencias probabilísticas. En ocasiones resulta difícil comprender el lenguaje que se utiliza y sobre todo la información científica de divulgación actual.

ABSTRACT

An important number of medical researches need the utilization of descriptive statistics that are going to describe the characteristics of the population we are studying. This description of variables depends in a fundamental way of the distribution of the variables. Following this, the description of a population with variables of scale of interval or variables with a scale of quotient will need the measures of central tendency and the measures of dispersion. Some of the measures of frequency of the disease more used are the ratio, the proportion or the rate.

Key words: Statistics, descriptive statistics, mean, median, standard deviation, ratio, proportion, rate.

Es importante para los médicos recordar los conceptos estadísticos y tratar de vencer la barrera profesional que tradicionalmente se ha presentado y que dificulta, en alguna medida, la interpretación de los resultados de la investigación médica.

En este artículo se presentan conceptos básicos de estadística, principios que ayudarán al médico clínico en la comprensión o realización de investigaciones.

La investigación clínica se apoya en el método estadístico para la recolección, organización, análisis e interpretación de una serie de datos en la obtención de inferencias de lo que ocurre en una población determinada, a partir de una muestra representativa de la misma.

Una cantidad importante de investigaciones requieren de estadística descriptiva. Las estadísticas que podemos obtener son de dos tipos:

a) Estimación puntual. Consiste únicamente en determinar un punto dentro de una escala numérica. Ejemplos: media, proporciones, diferencias entre dos promedios, diferencias entre dos proporcio-

¹ Jefa de la División de Investigación Epidemiológica. Hospital General "Dr. Manuel Gea González". Coordinación de Asesores, Subdirección General Médica, ISSSTE.

² Jefe de la División de Información y Comunicación Médica. Hospital General "Dr. Manuel Gea González".

³ Jefe de la División de Investigación Clínica. Hospital General "Dr. Manuel Gea González". Departamento de Cirugía, Facultad de Medicina, UNAM.

Correspondencia:

Dra. Pilar Mata Miranda.

Jefa de la División de Investigación Epidemiología. Dirección de Investigación. Hospital General "Dr. Manuel Gea González". Secretaría de Salud. Calzada de Talpan 4800, Col. Toriello Guerra, Delegación Tlalpan. 1400, México, D.F. Tel. 56 66 60 21

nes, varianza de una población o varianza de dos poblaciones.

- b) Estimación por intervalos.** Es la determinación de un intervalo de valores que incluirán el valor real de un parámetro de la población.

VARIABLES Y ESCALAS DE MEDICIÓN UTILIZADAS EN LA DESCRIPCIÓN DE LOS RESULTADOS DE UNA INVESTIGACIÓN

Las variables son las características o atributos que toman diferentes valores de un individuo a otro, de una población a otra, por ejemplo pH de la sangre, distensión abdominal, estatura, presión arterial.

Para describir una variable, se debe conocer su escala de medición. De acuerdo a la escala, las variables se clasifican en cuantitativas y cualitativas.

Las variables cuantitativas son de naturaleza numérica y se obtienen con instrumentos de medición (peso, talla, presión arterial, valores de hematócrito); pueden ser continuas, cuando no poseen interrupciones en la escala de valores que se presente (ejemplo: talla: 1.79 m; peso: 51.99 kg), o pueden ser discretas cuando tienen interrupciones en la escala de valores (ejemplo: número de hijos, ingresos en una guardia, defunciones hospitalarias).

La escala de razón, utilizada para variables cuantitativas, es el nivel más alto de medición; consiste en obtener los valores reales de cada medición individual. El cero es real. La escala de intervalos que se aplica a variables cuantitativas consiste en ordenar los valores de una serie de datos, estableciendo distancias definidas entre dichos datos.

Las variables cualitativas son características que no pueden ser medidas con instrumentos (como el color de ojos, la religión o el estado civil); se clasifican en nominales y ordinales. En la escala nominal se designan o nombran las observaciones en varias categorías como el grupo y Rh, religión, estado civil y algunas pueden ser dicotómicas, por ejemplo, aprobado sí o no, género masculino o femenino. La escala ordinal consiste en clasificar por grados de acuerdo con algún criterio, por ejemplo, vómito leve, moderado, severo; estado socioeconómico bajo, medio, alto.

Ejemplo de la utilización de diferentes tipos de escalas de las variables para un mismo fenómeno.

Evaluación de la diarrea.

Escala de razón: número de evacuaciones por día = 20.

Escala de intervalo: número de evacuaciones por día = de 1-5; 6-15, > 15.

Escala ordinal: diarrea leve, moderada, severa.

Escala nominal: con diarrea, sin diarrea.

MEDICIÓN CLÍNICA-DESCRIPTIVA

La estadística descriptiva tiene como objetivo mostrar el comportamiento en conjunto de los valores que se obtienen a través de la medición de las variables elegidas. Muestra los detalles simples de cada variable en el grupo de individuos que son estudiados así como su distribución.

Las medidas de resumen de las variables cuantitativas en escala de razón que se usan para la descripción son: media, mediana y moda. También se llaman medidas de tendencia central.

Las medidas de dispersión para variables cuantitativas en escala de razón son: varianza y desviación estándar. Indican qué tan alejados están los valores del centro de la distribución.

MEDIA

Es el promedio aritmético de las observaciones. Es el valor que representa a todos los individuos de la muestra. La fórmula es:

$$\bar{X} = \frac{\sum x}{n}$$

Donde x es el valor de cada una de las observaciones y n representa el total de las observaciones.

Los pasos para obtener la media son:

1. Sumar los valores de todas las observaciones.
2. El resultado de la suma se divide entre el número total de observaciones.

Ejemplo: Se pretende describir el promedio de edad de los 10 niños que ingresaron por primera vez a una guardería. Las edades registradas en años fueron: 1, 2, 3, 2, 1, 4, 1, 1, 3 y 4. La suma de las edades de todos los niños es 22, dividido entre los 10 niños es igual a 2.2 años. Por lo tanto, el promedio de edad de los niños que ingresaron por primera vez a una guardería es de 2.2 años.

Uno de los inconvenientes de esta medida de resumen es su sensibilidad a valores extremos, tanto muy altos como muy bajos, ya que la media se ve influida por estos valores. Cuando esto ocurre es preferible utilizar otra medida de resumen como la mediana.

MEDIANA

Es el valor que divide exactamente a la mitad una serie de valores ordenados de menor a mayor, lo que provee una lista de observaciones con los valores más bajos y otra de observaciones con los valores más altos. Cuando la cantidad de observaciones es impar, la mediana es el valor de la observación que justamente está en medio de la serie ordenada de valores; su fórmula es $\text{mediana} = \frac{n + 1}{2}$

Cuando el número de observaciones es par, la mediana es el promedio de los dos valores centrales, ordenados en forma progresiva.

Pasos para obtener la mediana:

1. Ordenar los valores de las observaciones de menor a mayor.
2. Identificar la ubicación del valor de la mediana.
3. Cuando la serie es par, identificar los dos valores centrales, sumarlos y promediarlos.

Ejemplo: Encontrar la mediana de la hemoglobina glicosilada (medida en %) de 14 personas con diabetes mellitus tipo II tratados con glibenclamida: 8, 11.9, 8.5, 7.4, 8, 9.6, 8.5, 6.7, 7, 10.4, 7, 8.2, 8.7, 8.1.

Observaciones ordenadas de menor a mayor:

6.7, 7, 7, 7.4, 8, 8, **8.1, 8.2**, 8.5, 8.5, 8.7, 9.6, 10.4 y 11.9. Se obtiene el promedio de los números centrales marcados en negritas $\frac{8.1 + 8.2}{2} = \frac{16.3}{2} = 8.15$, que es el valor de la mediana.

MODA

Es el valor que más se repite en una serie de datos. En algunos casos, es posible encontrar dos o más modas por lo que la distribución se considera bimodal o trimodal. En el ejemplo anterior hay tres modas, ya que los valores 7, 8 y 8.5 se repiten en dos ocasiones.

Si la media, la mediana y la moda adoptan el mismo valor en una serie de datos, podemos concluir que la distribución se acerca a la normalidad.

VARIANZA Y DESVIACIÓN ESTÁNDAR

La varianza es la dispersión que presentan los datos con respecto al valor promedio. Da cuenta de la variabilidad de los valores tomando como referencia al promedio.

La fórmula es:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (x - \bar{x})^2$$

Debido a que la varianza es muy compleja de interpretar, ya que el resultado se expresa al cuadrado, por ejemplo litro cuadrado, metro cuadrado, mmHg cuadrado, es conveniente utilizar una medida de dispersión más comprensible como lo es la desviación estándar.

La desviación estándar es la distancia a lo largo de un eje horizontal entre el promedio y el punto donde la curva pasa de convexa a cóncava y nos brinda información del porcentaje del área bajo la curva que se logra tener, es decir, el porcentaje de los valores de que se encuentran bajo la curva de distribución normal. Se consideran como normales los valores ubicados entre la media y más menos dos desviaciones estándar, que equivalen al 95.4% (Figura 1).

La desviación estándar se obtiene simplemente calculando la raíz cuadrada de la varianza y la fórmula es:

$$S = \sqrt{S^2}$$

Siguiendo el ejemplo anterior, el valor promedio de la hemoglobina glicosilada sería de 8.35 y la varianza de estos valores con respecto al promedio sería de 1.54 al cuadrado. Si calculamos la raíz cuadrada de 1.54 obtenemos la desviación estándar que sería en este caso de 1.24. Por lo anterior, los valores normales se encontrarían entre 8.35 más menos 2 (1.24), lo que da un intervalo de valores entre 5.87 y 10.83.

MEDIDAS DE FRECUENCIA DE LA ENFERMEDAD MÁS UTILIZADAS EN LA PRÁCTICA CLÍNICA

RAZÓN

Es la comparación, a través de una división, de dos grupos de individuos con atributos de diferente naturaleza. Es la forma más simple de mostrar desigualdades entre grupos.

Continuando con otro ejemplo hipotético, se han atendido a 1,200 pacientes con diabetes en el Servicio de Urgencias, 900 de ellos fueron mujeres y 300 hombres. En la división puede considerarse como numera-

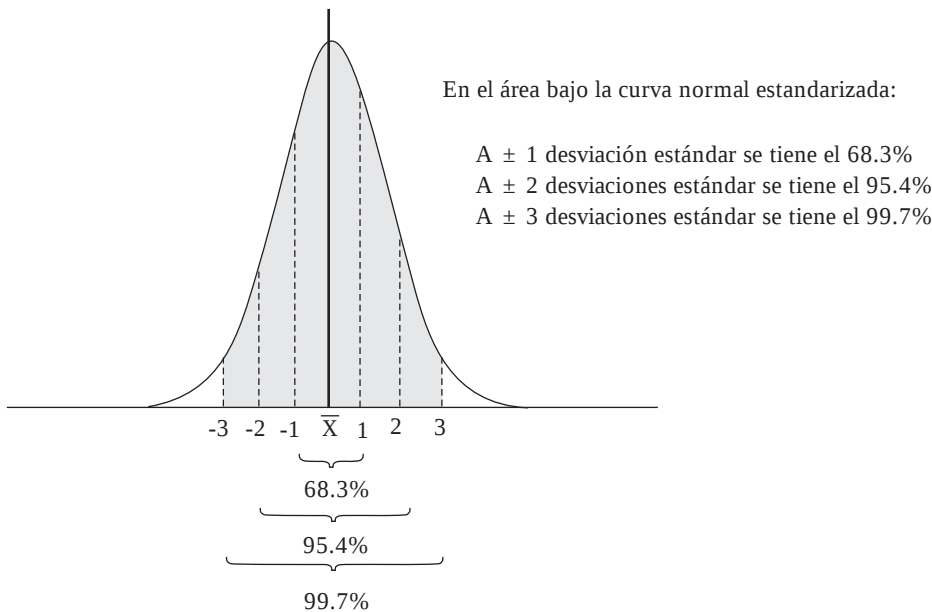


Figura 1. Curva de distribución normal.

dor cualquiera de los valores. De ahí que $900/300 = 3$, por lo que la relación entre mujeres y hombres con respecto a diabetes mellitus que acudieron al Servicio de Urgencias fue de 3 mujeres por cada hombre (3:1).

PROPORCIÓN

La proporción es la medida de estadística descriptiva que más se usa. Es el número de observaciones con una característica en particular entre la población de referencia. El numerador siempre está incluido en el denominador. Se expresa en porcentaje.

Las medidas de proporción utilizadas en la práctica clínica para describir la enfermedad son la prevalencia y la incidencia.

La prevalencia es el número de casos existentes de una enfermedad en particular entre la población de referencia.

Por ejemplo: la prevalencia de pacientes con cáncer de próstata diagnosticados en el Servicio de Urología en el 2004 sería:

Prevalencia = No. de pacientes con cáncer de próstata: 75

Total de pacientes del Servicio de Urología: 1,500

$75/1500 = 0.05$. Generalmente la fracción resultante se multiplica por 100, debido a que la probabilidad es de 0 a 100 y se expresa en porcentaje. Esto es, la prevalencia de pacientes con cáncer de próstata diagnosticados en el Servicio de Urología durante el 2004 fue de 5%; dicho de otra manera, aproximadamente 5

de cada 100 pacientes que acuden al Servicio de Urología son diagnosticados con cáncer de próstata.

La incidencia, a diferencia de la prevalencia, es el número de casos nuevos que se presentan de una enfermedad en particular dividida entre la población libre de la enfermedad al inicio del seguimiento.

Incidencia:
$$\frac{\text{Número de casos nuevos}}{\text{Población libre de la enfermedad al inicio del periodo de seguimiento}}$$

Ejemplo: Se pretende estimar la incidencia de infección de sitio quirúrgico en cirugías electivas en un hospital de segundo nivel. A lo largo del trimestre se presentaron 25 casos nuevos de infección de sitio quirúrgico y en ese mismo periodo se realizaron 789 cirugías.

Incidencia:

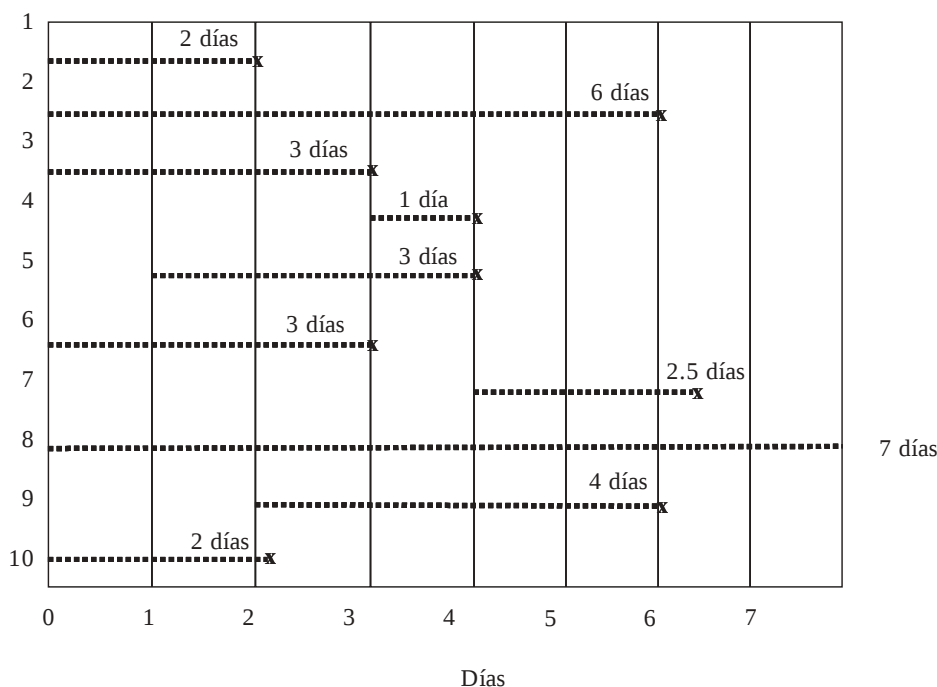
$$\frac{\text{No. de casos nuevos de infección de sitio quirúrgico}}{\text{Población expuesta a cirugía electiva}} = \frac{(25)}{(789)} \times 100 = 3.17\%$$

es decir, durante el trimestre, de cada 100 pacientes que se operaron, aproximadamente 3 desarrollaron infección de sitio quirúrgico.

TASA

La tasa es la medida de frecuencia que da cuenta de la velocidad de cambio en la población, de estar sana a

Paciente



..... Tiempo que transcurre antes de la recuperación (tiempo postoperatorio)

X Recuperación

Figura 2. Tasa de recuperación de pacientes postoperados de colecistectomía por cirugía de invasión mínima.

pasar al estado enfermo, o del estado de enfermo a sano. La tasa es la comparación, a través de una división, entre el número de veces que se presenta una enfermedad, o el evento estudiado, en un periodo determinado, entre el tiempo-persona o tiempo poblacional, que es la expresión con la que cada individuo contribuye libre de la enfermedad, o del evento, en el tiempo de seguimiento.

Por ejemplo, se pretende estimar la tasa de recuperación de los pacientes operados de colecistectomía por invasión mínima. Se siguió a 10 pacientes durante 7 días.

Se construye un cuadro con la evolución de los pacientes en días (Figura 2), se realiza la suma de los días que tardaron en recuperarse, se divide el número de pacientes recuperados (9 pacientes) entre la sumatoria de los días-persona. La fracción se multiplica por 1,000 días persona, ya que la tasa se expresa en múltiplos de mil.

Tasa de recuperación = $9/33.5$ días persona = $0.269 \times 1,000$ días persona = 269. La interpretación es que si se siguiera a 1,000 pacientes postoperados de colecistectomía con cirugía de invasión mínima en un día, se recuperarían 269. La tasa debe describirse de

acuerdo a la unidad de tiempo que puede ser día, mes o año. El paciente número 8 no se recuperó, pero se contaron los 7 días de contribución sin recuperación.

CONCLUSIÓN

La estadística descriptiva es en sí misma una herramienta fundamental para el análisis de datos de los estudios correspondientes. Los conceptos que se han presentado en este artículo son los básicos y es importante que los investigadores clínicos profundicen en el tema para lograr un mejor entendimiento de los fenómenos en el campo de la medicina.

REFERENCIAS

1. Saket DL, Haynes RB, Guyat GH, Tugwell P. Epidemiología clínica. *Ciencia básica para la medicina clínica*. 2a edición 3ª impresión. México: Editorial Médica Panamericana, 1998.
2. Pagano M, Gauvreau K. *Fundamentos de bioestadística*. 2ª edición. México: Math Learning, 2001.
3. Silva LA. *Cultura estadística e investigación científica en el campo de la salud: una mirada crítica*. Ediciones Díaz de Santos, 1997.
4. Dawson-Saunders B, Trapo RG. *Bioestadística médica*. 2ª edición. México: Manual Moderno, 1997.