AUTORA: DRA. SILVIA BEATRIZ D'ANGELO

MHS Medicina Hombre y Sociedad
Epistemología e introducción a la investigación
FACULTAD DE MEDICINA

Conceptos básicos de estadística aplicada ala investigación científica.

Volumen I, n° I 2022

Contenido:

Introducción a la Estadística. Glosario de conceptos básicos 1
Presentación de datos. 2
Medidas de resumen para datos cuantitativos. 3
Medidas de resumen para datos cualitativos. 4
Medidas de asociación Significación estadística 5
Bibliografía 6

Introducción a la Estadística.

La **Estadística** es una ciencia que brinda instrumentos e ideas para la recopilación , organización, presentación, análisis e interpretación de datos cuantitativos, obtenidos de un conjunto de individuos, grupos o serie de hechos, sujetos a variabilidad aleatoria.
Tiene por **objetivo** deducir las leyes que rigen los fenómenos en estudio, y sacar concusiones que permitan predecir hechos futuros y tomar decisiones.
Se clasifica en **Estadística descriptiva** mediante la cual se detallan y caracteriza un grupo de datos.
Estadística inferencial, permite conocer la población de estudio a partir de conocer una parte de ella.
La **utilidad** de la Estadística esta asociada al a investigación cuantitativa, su conocimiento



permite la lectura e interpretación de informes estadísticos e informes científicos, que permiten una practica profesional, basada en información actualizada y válida.

Glosario de conceptos básicos

Población: Conjunto de unidades elementales que poseen características que el investigador desea estudiar. Puede estar constituido por individuos, objetos, instituciones, animales, registros médicos, fenómenos como accidentes viales, otros. Pueden ser finitas, cuando hay un marco muestral en el cual encontrar todos los elementos de la población, o infinitas

Muestra: Subconjunto de la población que posee las características que se desean conocer. Pueden ser probabilísticas o no probabilísticas. *Las muestras probabilísticas,* donde cada unidad que conforma la población tiene la misma posibilidad de ser elegida, son muestras que utilizan el azar para seleccionar las unidades muestrales, y permiten generalizar los datos a toda la población. Se encuentran dentro de este tipo de muestras las aleatorias simples (o por sorteo), aleatorias sistemáticas, por conglomerados y por estratos.

Estas son las muestras más representativas, presentando las características de la población tanto en calidad como en cantidad.



De acuerdo al tipo de variable que se estudien, se seleccionarán los métodos estadísticos para la presentación y el análisis de los datos.



Dentro de las muestras no probabilísticas tenemos las muestras por conveniencia, accidentales y voluntarias, entre otras. Sus resultados solo son referidos a los sujetos que tengan las mismas características que los estudiados.

Estadístico: dato que describe a la muestra.

Parámetro: Dato que describe a la población.

Variables: características o atributo a estudiar en una o varias unidades elementales.

Pueden ser cualitativas o cuantitativas.

Las cualitativas se expresan con palabras. Ejemplo: el nombre de una persona. Se subclasifican en dicotómicas cuando admiten dos valores, por ejemplo el sexo de un apersona: Femenino o Masculino. También pueden ser policotómicas, cuando tiene más de dos valores. Ejemplo: Variable nacionalidad: Argentino, Frances, Español, otras. A su vez las variables cualitativas, según la escala de medición , pueden ser: nominales cuando sólo se listan las categorías asumidas: Ejemplo: Variable: Tipo de alimento según su composición química. Valores: Glucídicos, proteicos y lipídicos.

Cuando los valores que asume la variable presentan un orden se denominan ordinales. Ejemplo: Variable: Niveles de atención a la salud. Valores: Primer nivel, segundo nivel , tercer nivel.

Las variables cuantitativas se expresan numéricamente y se sub– clasifican en discretas y continuas. Las discretas no admiten valores intermedios entre una variable y otra. Ejemplo. Cantidad de piezas dentales. Las variables continuas aceptan valores intermedios. Ejemplo peso de una persona. Entre 60 kg. Y 61 kg. Podemos tener otros valores (60,200 kg– 60,330 kg, 60,500 kg y así hasta llegar al 61 kg)

Se miden en escalas de razón o proporción o escalas de intervalo

Presentación de datos

Los datos pueden presentarse mediante textos, cuadros y tablas, estadísticas y gráficos estadísticos.

El TEXTO es forma expositiva de presentación que describe los objetivos y el proceso de construcción y análisis del dato y las conclusiones a las que se han abordado. Permite resaltar datos, detallar el método, obtener información adicional y detallada de los datos resaltando situaciones especiales encontradas en el estudio. Su mayor desventaja es que si la información es mucha puede tornarse confuso.

CUADROS Y TABLAS: Este tipo de presentación utiliza una grilla colocando en la primer columna y en la primer fila las variables y sus categorías, y en el las demás celdas las frecuencias con que se presenta cada categoría. A diferencia de las tabla los cuadros llevan totales y subtotales en la última columna y en la última fila. Es una forma de presentación clara y concisa, permitiendo una rápida comparaciones de los datos que presenta. Su desventaja es que tiene menos información que el texto y puede ser muy extenso dependiendo de la cantidad de datos.

GRÁFICOS: Es una forma de presentación de datos estadísticos mediante un dibujo que resume las principales características cuantitativas del conjunto de observaciones medidas y sirve para comparar la intensidad de las variables y de las categorías de estudio. La presentación grafica no pretende brindar precisión en los valores de las categorías que se presentan; más bien, es un auxilio visual para facilitar la comprensión de información numérica.

Para variables cualitativas y cuantitativas discretas se utilizan gráficos de barras simples para presentar una variable, barras agrupadas para presentar una variable en dos poblaciones o situaciones diferentes y gráficos de sectores. Para variables cuantitativas continuas, histogramas y polígonos de frecuencia, para mostrar la evolución de una variable en el tiempo se utilizan gráficos lineales, para compara dos variables cuantitativas continuas lo gráficos de correlación. También tenemos los cartogramas muy útiles en epidemiologia para ver distribución de enfermedades.

Para leer un gráfico recordemos que los gráfico de barra al igual que los lineales se construyen sobre ejes cartesianos. En la abscisa (eje x) va la variable y en la ordenada (eje y) la frecuencia con la que se presenta la situación observada. En los gráficos de sectores cada sector representa las observaciones realizadas por cada categoría de la variable. Sólo los gráficos correlacionales presentan dos variables, una en cada eje, mostrando su relación , por ej, edad en el eje "x" y talla en el eje "y"

Los gráficos dan una idea rápida de la situación general. Su mayor desventaja es que puede distorsionarse la información si no se manejan bien las escalas o los colores.



Los datos cuantitativos se describen siempre con dos medidas las MTC Y su correspondiente MD.

Medidas de resumen para datos cuantitativos

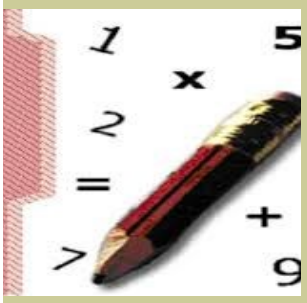
Los datos numéricos—o cuantitativos– se pueden sintetizar según sea el objetivo de análisis con las siguientes medidas:

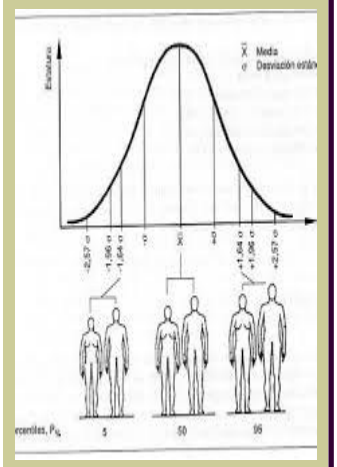
MEDIDAS DE TENDENCIA CENTRAL: Las MTC tienden a ubicarse en el centro de una serie de observaciones y sirven como valores representativos de la misma. Brindan información sobre el punto (o tramo) de la escala de referencia donde se posiciona la distribución.

Son MTC la media aritmética que por ser sensible a valores extremos se recomienda usarla cuando la serie de datos es homogénea. La mediana que siempre se ubica en el medio y se usa cuando la serie es heterogénea. Y la moda que es el valor que mas se repite.

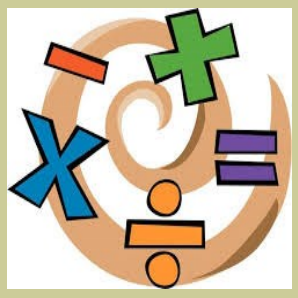
MEDIDAS DE POSICION: Fraccionan la serie en partes iguales (cuartiles, quintiles, deciles y percentiles) y se utilizan cuando se quieren analizar diferentes sectores de la serie de observaciones realizadas.

Los cuartiles dividen la serie en cuatro partes iguales. Cada cuartil representa el 25% de la serie.





Cuando las tres MTC coinciden en un mismo punto, se dice que la distribución de la serie de datos , es normal. Gráficamente tiene forma de una campana simétrica



Los quintiles dividen la serie en cinco parte , cada quintil equivale al 20 % de la serie, Los deciles la dividen en 10 partes iguales y cada decil equivale al 10% de la serie y los percentiles dividen la serie en 100 partes iguales equivalentes cada una al 1% de la serie. Esta medida se utiliza cuando las observaciones son superiores a 100, algunos autores dicen que son útiles cuando se tiene alrededor de 300 observaciones .

MEDIDAS DE DISPERCIÓN: Las MD son MR que sirven para expresar como se distribuyen las observaciones en una serie. Son el complemento necesario de las MTC para describir una distribución de frecuencias. Son MD: desvió estándar (DS), desvió intercuartilico, (DI) y el rango o amplitud

Las MTC indican en que punto de la escala se ubica la distribución y las MD como se dispersan sus observaciones. Cuanto más chico es valor de la MD , más agrupados alrededor de la MTC están los datos de la serie. Cuanto má grande la MD , más variabilidad hay en la serie de datos. EJ. Edad promedio (μ) 20 años ± 2 DS quiere decir que en las observaciones hay edades entre 18 y 22 años . Si fuera Edad promedio (μ) 20 años ± 4 DS la s edades en que se encontrarían las observaciones alrededor de la media sería mín. 16 y máx.24, lo que nos muestra una serie más dispersa.

El DS acompaña siempre a la media aritmética o promedio. EL DI acompaña a la mediana y el rango amplitud acompaña al modo.

Medidas de resumen para datos cualitativos

Los datos obtenidos de variables cualitativas como el nivel de instrucción alcanzado se sintetizan mediante CIFRAS RELATIVAS. Estas cifras son cocientes que relacionan dos valores absolutos.

TASAS: una tasa relaciona un fenómeno ocurrido en un lugar (I) y un periodo de tiempo (t) determinado con la población expuesta a ese fenómeno. El fenómeno a expresar se coloca en el numerador y la base de comparación en el denominador. Los casos del fenómeno deben provenir de la población que se define como expuesta; y el lugar y el periodo de referencia deben ser iguales para el fenómeno y la población.. La tasa indica que el fenómeno tenemos en el numerador (ej, mortalidad) puede ocurrir en la población que tenemos en el denominador, en la proporción que indica la tasa. Ej, 17 por mil. Esto indica que en esa población pueden ocurrir 17 defunciones por cada 1000 habitantes.

La tasa indica el riesgo de que el fenómeno estudiado, se de en la población.

RAZONES: en las razones se relaciona dos fenómenos (variables) diferentes o dos características distintas de un mismo fenómeno. Por ej. Número de habitantes en un territorio / superficie de ese territorio. Por ej. Corrientes tiene 11 habitantes por km^2

PROPORCIONES: el numerador forma parte del denominador y da una idea del peso que tiene una parte respecto del todo. Es la expresión de la participación de una categoría en el total de la variable de referencia. Por ej. Si calculamos la cantidad de población de hasta 25 años en el total de la población de un país.

Medidas de asociación.

En el estudio de asociación entre variables se analizan además de la existencia de la asociación, la dirección, la naturaleza y la intensidad.

En cuanto ala dirección la asociación puede es directa o indirecta . Una asociación directa es cuando al aumentar el valor de una variable, aumenta el valor de la otra. Por ejemplo: Aumenta la estatura y aumenta el peso de una persona. Una asociación negativa se da cuando al aumentar el valor de una variable la otra disminuye. Por ejemplo: Se aumenta la dosis de un ansiolítico disminuye el nivel de ansiedad de la persona.

Existen diferentes medidas para valorar la asociación entre dos variables. Una de esas medidas cuando las variables son cuantitativas es el coeficiente de correlación lineal "r" . Dado que las asociaciones pueden ser directa (o positivas) o indirectas (o negativas), este coeficiente puede asumir valores que van de –1 a +1. Si "r" es igual a 0 no existe correlación, si los valores están entre –25 a +25, la asociación es débil; si el resulta esta entre –50 y +50 la asociación es moderada, si el valor s encuentra entre –75 y +75, la asociación es fuerte y si el valor de "r" es igual a –1 o +1 la asociación es perfecta.

Cuando las variables son cualitativas una de las medidas a utilizar es el coeficiente de contingencia cuyos valores oscilan entre 0 y 1.

Otro coeficiente utilizado es el coeficiente de Ro, que mide la concordancia entre dos variables. Este coeficiente también puede de tomar valores comprendidos entre -1 y 1

Al hablar de asociación entre variables se asume que una variable influye sobre otra directa o indirectamente. La variable que influye sobre otra modificando el comportamiento de la segunda se denomina variable independiente (VI). La variable que es modificada por acción de otra se denomina variable dependiente (VD).

En epidemiología la asociación estadística entre un facto causal (VI) y un efecto (VD) n e estudio de asociación dese calcula con el Riesgo Relativo (RR) o la Razón de Productos cursados (OR). Cuando el valor que se obtiene es mayor a 1 indica que la población expuesta a determinado factor tiene tantas probabilidades de desarrollar determinado efecto, como sea el valor obtenido.

Significación estadística.

La pruebas de significación estadística miden la probabilidad influencia del azar en los resultaos obtenidos al realizar la medición de una variable, en una muestra.

Esta probabilidad se representa con la letra "p" y se calcula con diferentes pruebas como Chi 2 y la prueba "t" de Student.

Cuando "P" es igual o menor de 0,05 se dice que los resultados son significativos.

Este es un valor asumido por convención de expertos y quiere decir que los resultados se pueden deber al azar sólo en un 5% o menos.



Una asociación directa es aquella donde al aumentar el valor de una variable, aumenta el valor de la otra. Una asociación indirecta es cuando al aumentar el valor de una variable, disminuye el valor de la otra.



Este material se encuentra en proceso de revisión. Prohibida su reproducción. Por comentarios o sugerencia dirigirse a la dirección electrónica de la autora.

Autora: SILVIA BEATRIZ D'ANGELO.
Doctora en Ciencias de la Salud .
Mgter y Esp. en Docencia Universitaria.
Especialista en Salud Pública.
Licenciada en Kinesiología y Fisiatría.
Año: 2022
Correos de contacto:
doctorasbdangelo@gmail.com
silviabdangelo@hotmail.com



Este material ha sido elaborado para la asignatura Medicina, Hombre y Sociedad, del primer año de la carrera de Medicina de la Facultad de Medicina perteneciente a la Universidad Nacional del Nordeste.

En el mismos se realiza una compilación de diferentes fuentes bibliográficas cuya finalidad es acercar al estudiante conceptos básicos, necesarios para:

-El reconocimiento de información científica dentro de la literatura médica —Diferenciación entre información científica y no científica. – Selección de información válida al momento de realizar una búsqueda bibliográfica

Bibliografía

DAWSON Y TRAPP- (2005) *Bioestadística Médica* 4ta. Edición. Edit El Manual Moderno. S. A. Buenos Aires. Argentina .

HERNADEZ SAMPIERI y Col. (2011) *Metodología de la investigación* 5ta edición . Edit. Mc. Greaw Hill.

GUERRERO Y col (1986) Epidemiología. Serie Paltex Edit. Addison – Wesley Iberoamericana.

MARDER, G. (2003) Bioestadística Asociación Cooperadora, Facultad de Ciencias Veterinarias, Universidad Nacional del Nordeste.

SPIEGEL, M. (1997) Estadística. 2da.edic. . Edit. Mc Graw-Hill. España.