

BIOESTADÍSTICA:

La palabra estadística proviene del latín: "status": modo de pararse, posición.

En el lenguaje cotidiano se habla de estadística en dos sentidos, uno es para referirse a un conjunto determinado de datos, por ejemplo, a la estadística de población o estadística de ventas.

El otro sentido se refiere a una disciplina matemática, donde la estadística matemática es una de las áreas de la ciencia matemática.

Al mismo tiempo la estadística se estudia desde el punto de vista práctico, desde su aplicación, por ejemplo, la estadística aplicada, es su uso en la ciencia biológica y las disciplinas relacionadas con la medicina y la salud. A esto se le llama bioestadística.

La estadística es un conjunto de técnicas para el análisis de los datos. De esto se desprenden dos componentes principales de esta disciplina: DATOS y ANÁLISIS.

Datos:

Los datos son la materia prima de la estadística. Los datos salen de las mediciones o también llamados observaciones.

Al mismo tiempo, las técnicas de análisis estadísticas permiten que los datos se conviertan en información útil.

Según Mainland, la estadística es el método científico que se ocupa del estudio de la variación; y según Cobb y cols., la variabilidad es quizás lo único constante en nuestro mundo.

La estadística de salud con todos aquellos datos numéricos debidamente capturados, validados, elaborados, analizados e interpretados que se requieren para las acciones de salud.

La estadística en salud depende de los datos y de su análisis, por lo tanto, es esencial conocer la metodología para la correcta recolección de los datos y comprender la estadística para un correcto análisis de éstos.

Conceptos básicos:

Parámetro:

Cualquier número resumen de los elementos básicos de la biostatística de una población. otra definición de Villaruel 2018 el parámetro es una función de los datos calculada en la población.

Estadísticos:

Es cualquier número resumen de una muestra. También se le puede llamar parámetro estimado, que es el término correcto cuando se habla de un número resumen que deriva de una muestra.

Como es solo un estimador del parámetro, se denota con letras en latín; \bar{X} es la medida de una muestra; S^2 es la varianza.

Observar:

Es medir. lo observado es lo medido. las observaciones son las mediciones que se realizan en las unidades muestrales.

Unidad de análisis o unidad muestral:

Es el objeto con la característica de interés que será observado o que será medido, pueden ser personas, cortes histológicos, crecimiento de colonias, entre otros.

Atributo:

Característica de la unidad de análisis que se va observar.

Variable:

Cualquier característica que tome dos o más valores de una población.

Variable aleatoria:

Son aquellos atributos que han sido medidos en un conjunto de individuos de la población que conforman una muestra aleatoria y que no se puede anticipar el resultado.

Planificación estadística.

Cuando se escribe un proyecto de investigación, se diseña la metodología del estudio. y dentro de esta, es indispensable diseñar la planificación estadística (PE). Debe estar enfocada en cumplir cada uno de los objetivos específicos y el objetivo general y con ello responder la pregunta de investigación.

La PE consiste en:

- I: Definir y diseñar el método de selección de la muestra, esto se llama diseño muestral.
- II: Definir el número de elementos que conformara la muestra, esto se llama cálculo de tamaño muestral.
- III: Definir variables: El tipo de variables, la escala de medición y la unidad de medida a las que aplica.
- IV: plan de tabulación de datos descripción y análisis de la información recolectada. se define las formas de presentar y analizar la información recolectada.

I: Definir y diseñar el método de selección de la muestra, esto se llama Diseño muestral:

Concepto de estudio o blanco de muestra.

La población así también llamada universo de estudio, es el conjunto total de personas u objetos que tienen una característica en común de interés para un estudio.

MUESTRA: Es un subconjunto de individuos o elementos de una población definida que cumple con ciertas propiedades comunes.

I: Representativa de la población de estudio o población blanco:

Implica, que para que los resultados del estudio tengan validez interna y externa, la muestra debe ser representativa de la población blanco.

Si la muestra no es representativa de la población de donde procede, todos los cálculos que se hagan serán válidos solo para la muestra, sin posibilidad de extrapolar estos resultados a los individuos que no fueron incluidos en ella.

2- Sección aleatoria: Significa que los sujetos de la población blanco deben ser escogidos al azar, es decir, todos los individuos o elementos de la población blanco tienen que tener la misma probabilidad de ser seleccionados en la muestra.

3- Tamaño mínimo de acuerdo: Es el número de individuos necesarios y seleccionados aleatoriamente, que debe contener la muestra, para obtener estimaciones que representen los valores reales del parámetro de la población de estudio. Es por esto, que se debe realizar un cálculo del tamaño muestral.

1- DISEÑO MUESTRAL:

La selección de la muestra aleatoria es indispensable que para la obtención de resultados válidos.

MUESTRA ALEATORIA: Es una selección al azar de los individuos que componen la muestra. Existen diferentes tipos de métodos de selección de la muestra.

1.1: Muestreo probabilístico:

1.1.1: Aleatorio simple: Se asume que la población es homogénea y que todos los elementos de la población tienen la misma probabilidad de ser elegidos en la muestra.

Ventajas: Técnica sencilla

Desventajas: la muestra puede quedar desequilibrada.

1.1.2- Muestreo estratificado. Este diseño muestral tiene por objetivo evitar que, por azar, algún grupo esté menos representado que otro.

Se asume que la población es heterogénea y que existen subgrupos, y es por eso, que se agrupa la población en unidades homogéneas que se llaman estratos.

Ventajas: previene que la muestra quede desequilibrada.

Disminuye la variabilidad dentro de los estratos.

Desventajas: Complica un poco más el diseño muestral.

No se aconseja más de 8-10 estratos.

1.1.3- Muestreo sistemático: Se usa cuando los elementos de la población están ordenados. En este caso se elige al primer individuo al azar y el resto viene condicionado por aquel.

Ventajas:

- Obtiene buenas propiedades de representatividad, similares a la del muestreo aleatorio.

Desventajas:

Existe la posibilidad de que el orden en que se han listado los candidatos a la muestra tenga algún tipo de periodicidad oculta que coincida con el intervalo escogido para generar la muestra sistemática.

1.1.4: Muestreo por conglomerado: Se usa cuando el muestreo aleatorio simple es demasiado caro, por la gran magnitud de población y tampoco se tiene un listado de los individuos que la componen.

Ventajas:

- Útil para estudios epidemiológicos que desea abarcar poblaciones de diferentes regiones
- Útil para grandes "n" muestrales.

Desventajas:

- Exige un conocimiento previo de las zonas de estudio.

1.2- MUESTREO PROBABILÍSTICO

1.2.1- NO probabilístico o sin asignación aleatoria: Método de muestreo que no se basa en probabilidades, es un muestreo por conveniencia o de cuotas.

1.2.2: De asignación aleatoria: A pesar de que la muestra no tuvo selección aleatoria, la asignación de intervención a través de métodos aleatorios. Esto permite que los grupos receptores de las exposiciones sean los más similares posibles y que sean comparables.

Ventajas:

- Ventaja principal es la conveniencia para el acceso de las unidades muestrales.
- Simple, económico, rápido

Desventajas:

- El principal defecto, la falta de representatividad y validez externa.
- Imposibilidad de hacer generalizaciones y aseveraciones estadísticas.

11.- Definir el número de elementos que conformará la muestra, Cálculo de tamaño muestral.

Otra de las exigencias que debe tener una muestra es que tenga el tamaño adecuado para que represente cuantitativamente a la población de estudio, o blanco.

En investigación clínica se realiza cálculo de tamaño muestral (Ct_{mu}) para varios objetivos y existen diferentes métodos, según el objetivo del estudio.

2.1.- ESTIMAR EL PARÁMETRO

El objetivo del Ct_{mu} es obtener la estimación de un parámetro, o sea obtener un estadígrafo representativo de la población de estudio y que responde la siguiente pregunta: ¿cuántas unidades de análisis es necesario estudiar para poder estimar el valor de una variable con el grado de confianza deseado y que represente al parámetro de la población de estudio?

2.1.1.- Variabilidad del parámetro o probabilidad del evento que se desea estimar.

Cuando se desea estimar una media, se necesita la varianza, o mayor variabilidad tiene un atributo, se necesitará un mayor número de medidas, por lo tanto, una muestra más grande.

2.1.2.- Precisión de la estimación del "parámetro"

Es la amplitud del intervalo de confianza. Cuanto más precisa sea la estimación, más estrecho deberá ser el intervalo y más sujetos deberán ser estudiados.

2.1.3.- Nivel de confianza

por convención se fija en 95%. Corresponde a un valor de error α de un 5%. en que el "parámetro estimado" se equivoque por azar solo en un 5%.

* Ct_{mu} para estimar una media:

$$n = Z^2 \sigma^2 / I^2 \text{ Aquí se asume una distribución Gaussiana.}$$

* Ct_{mu} para estimar una proporción:

$$n = Z^2 (Pq) / I^2 \text{ Aquí se asume una distribución Bernoulli.}$$

2.- Para comparar proporciones.

El objetivo del Ctmu es obtener el tamaño muestral necesario que permite detectar estadísticamente una magnitud de diferencia entre 2 grupos, en el supuesto que realmente exista.

En toda investigación que desea comprobar una hipótesis, es necesario evitar los errores en el número y calidad de las mediciones.

Es decir, conseguir la máxima exactitud al medir, lo significa, procurar que exista validez en la medición y prevenir el error aleatorio.

2.1. Error tipo 1 o α :

Es el error aleatorio y que se refleja en el error tipo 1 o α . Éste se debe fijar a priori cuando se realiza un cálculo de tamaño muestral. ¿Cuánto error aleatorio el investigador está dispuesto a aceptar?

Se define como el error que comete el investigador al rechazar la hipótesis nula (H_0), siendo esta verdadera, dicho de otra forma, es rechazar la hipótesis nula (H_0) cuando es cierta.

VALOR P:

Al fijar el error α , se fija el valor P. Porque el valor P es el valor de la probabilidad de cometer error α , es decir cometer error aleatorio. El valor P es la probabilidad de "caer" en la zona de rechazo: esto es descartar H_0 siendo cierta. Para decirlo en términos más simples, hasta que el punto del resultado observado, es decir medido, es probabilísticamente compatible con la hipótesis planteada, o sea H_1 .

El valor P es equivalente al valor de significancia α , con la diferencia de que el valor P se calcula a partir de una muestra a posteriori, mientras que el valor α se fija antes de ejecutar el estudio a priori, cuando se realiza el cálculo de tamaño muestral. Por lo tanto, el valor P no tiene relevancia si previamente no se ha realizado un cálculo de tamaño muestral, donde se fija a priori el valor α y la magnitud de diferencia que se desea estadísticamente probar.

2.2: Error tipo II o β

Es aceptar H_0 cuando no es cierta. Concluir que no existe asociación entre una variable independiente (X) y una variable dependiente (Y) cuando realmente sí existe. Se define como el error que comete el investigador al no rechazar la hipótesis nula, siendo esta falsa.

La magnitud del error β es otro de los supuestos que se debe fijar para el cálculo de tamaño muestral.

2.3: Magnitud de diferencia.

La magnitud de la diferencia que se postula encontrar tiene una gran influencia en el tamaño de la muestra. Se debe definir la mínima magnitud de diferencia que se desea detectar, y que sea de relevancia clínica para el paciente, por lo tanto, debe fijarse en términos realistas.

2.4: Tipo de hipótesis: Si es bilateral o a dos colas, en que postula que existe diferencia o que no exista diferencia.

Si es unilateral o una cola, en que postula que existe una diferencia mayor o menor que.

2.5: La probabilidad del efecto estándar (P_1) y la probabilidad del efecto que se propone (P_2).

Dependiendo del error alfa y el tipo de hipótesis, derivan el valor P .

¿Que implica un cálculo de tamaño muestral incorrecto?

a) Un exceso de tamaño muestral: Encarce el estudio. No es ético someter a más pacientes de lo necesario a la intervención o la vulneración de los datos.

b) Un insuficiente tamaño muestral, el "parámetro estimado" será poco o preciso.

Es correcto mencionar que existen varios métodos para el cálculo de tamaño muestral, según el diseño de estudio son: estudios de correlación, estudios de pruebas diagnósticas, estudios de no inferioridad, estudios caso control.

III: DEFINICIÓN DE LAS VARIABLES : TIPO VARIABLES, LA ESCALA DE MEDICIÓN Y LA UNIDAD DE MEDIDA.

parte de la planificación metodológica de un proyecto de investigación es definir las variables independientes, o predictivas, la variable dependiente o "outcome" y las co-variables. Durante la planificación estadística también se deben definir las variables, respecto al tipo, su unidad de medida y como se describirá según el tipo de medición que tiene la variable.

Según la escala de medición existen diferentes tipos de variables:

- Las que miden una cualidad (nominal o categóricas o las ordinal) se llaman CUALITATIVAS.
- Las que miden cantidad, se llaman CANTITATIVAS, que pueden ser discretos o continuos.

1: CUALITATIVAS:

1.1: Nominal o categóricas: Son aquellos que se clasifican en categorías o etiquetas. Dentro de éstas pueden ser dicotómicas, en que solo hay dos categorías y son auto-excluyentes.

Policotómicas, hay más de dos categorías y no necesariamente son auto-excluyentes.

En las variables nominales no hay un orden establecido, no tienen un valor:

1.2: Ordinal: Tiene un sentido de orden, que está implícito, sin que sea una magnitud o cantidad. Por lo que no tienen unidad de medida. No obstante, existen instrumentos que se usan para evaluar una condición, donde el autor le asigna un número para facilitar la interpretación de éste, pero eso no significa que la cualidad sea una cantidad. Un ejemplo de esto son los pruebas de evaluación de funcionalidad que tienen puntaje, que son construcciones artificiales donde una cualidad y según su nivel de cumplimiento, se le asigna un valor numérico arbitrario.

Ejemplos: Niveles de gravedad: (Leve, moderada severa.)

Escala o puntajes: EVA (1-10), Glasgow (3-15), Apgar: (1-9).

2- CUANTITATIVAS.

Son variables que son una cantidad, se refleja una magnitud, por lo que existe un orden natural en estas variables en la escala numérica. Tienen una unidad de medida.

Existen 2 tipos de variables cuantitativas:

2.1: Discreta: Atributo que no puede tomar valores decimales. Son valores de conteo, tales como, días de hospitalización, n° de hijos, n° de fallas, n° de muertes.

2.2: Continua: Una variable continua es aquel atributo que puede asumir un número infinito de valores dentro de un determinado rango. Tienen una unidad de medida y tienen una distribución en el plano cartesiano del eje x, también llamadas intervalares. Las variables medidas en escala de razón, el cero indica la ausencia de la variable.

Distribución Normal, Paramétrica o Gaussiana $\approx N$:

Fue reconocida por primera vez por el francés Abraham de Moivre. Posteriormente, Carl Friedrich Gauss formuló la ecuación de la curva de normalidad "la campana de Gauss". La distribución de una variable normal está completamente determinada por dos parámetros, su media y su desviación estándar.

IV: PLAN DE TABULACIÓN DE LA BASE DE DATOS, DESCRIPCIÓN Y ANÁLISIS DE LA INFORMACIÓN RECOLECTADA.

1- Plan de tabulación de la base de datos:

Una base de datos es una estructura en la cual se almacenan, con un orden definido, un grupo de descripciones sobre determinado sujeto o unidades de análisis y a medidas (variables), para finalmente ser un conjunto de datos pertenecientes a un mismo contexto y almacenados sistemáticamente para su posterior uso.

1- ESTADIGRAFOS DE ORDEN:

- El máximo $X(n)$ y el mínimo, $X(1)$ aparecen en forma instantánea al ordenar la muestra.
- PERCENTILES: A cada uno de los números que dividen la muestra en 100 partes iguales, en consecuencia, ellos son 99, y se denotan por $P(k)$, donde k es el orden del percentil indicado. Los percentiles más conocidos y usados:
 - Los cuartiles: son tres denotados por $Q_1, Q_2, y Q_3$, que corresponden respectivamente a los percentiles $P_{25}, P_{50}, y P_{75}$, ellos dividen la muestra en cuatro partes iguales

2- Estadígrafos de centralización o de tendencia central.

Cuando se observa un fenómeno cuantitativo, interesa saber si los datos recolectados se aglutinan en torno en ciertos valores representativos que son propios del fenómeno estudiado.

Son:

MODA: Dato de mayor frecuencia de aparición.

MEDIANA (P_{50}): Punto que divide a la muestra en 2 partes iguales.

MEDIA O PROMEDIO: Es el punto donde se ubica el centro de las masas de la muestra.

ANÁLISIS DE DATOS:

La estadística analítica tiene por objetivo comprobar hipótesis o establecer relaciones de causalidad en un determinado fenómeno. Es también llamada "Estadística Inferencial".

Para ello el análisis estadístico se podría ordenar en tres etapas:

1° Comparación del efecto entre grupos que están y no están expuestos a la variable independiente.

para ello se utilizan pruebas estadísticas de comparación y según la distribución de las variables, serán pruebas paramétricas (si la variable distribuye normal) o no paramétricas (si la variable no distribuye normal). Según la escala de variable que se compara.

La definición de las variables tanto en su tipo (Independiente, dependiente y co-variables) como en su escala de medición, es el paso previo básica para poder construir la base de datos. Aquellos que serán dicotómicos, los que serán policotómicos o tendrán categorías o cuales variables se registrarán en forma numérica.

Protección de la confidencialidad de la Información

Es importante que los Investigadores tengan presente que todo dato es del sujeto de estudio y por lo tanto, es necesario solicitar su consentimiento informado para registrarlos, describirlos y analizarlos.

Toda persona tiene derecho a la protección de la confidencialidad de su información y es por ello, que los análisis estadísticos deben realizarse sobre bases de datos codificados, donde no exista información que pueda identificar al sujeto de estudio.

Descripción de los datos.

La estadística descriptiva tiene el objetivo de describir cuantitativamente en conjunto de datos.

Para ello se utilizan diferentes recursos estadísticos, tales como los estadígrafos de orden, centralización y variabilidad, que son números resúmenes de los datos recolectados, que pueden provenir de estudios poblacionales o muestrales.

Estadígrafos de orden, centralización y variabilidad.

La estadística llama estadígrafos o estadísticos, a números resúmenes que permitan extraer conclusiones acerca de la estructura de una muestra o una colección de datos.

Se construyen estadígrafos para distintos fines. Los 3 tipos más conocidos por su amplio uso en la estadística descriptiva son:

- 1: Estadígrafos de orden
- 2: Estadígrafos de tendencia central
- 3: Estadígrafos de variabilidad.

Variables Categricos: Se usa la prueba exacta de Fischer, en caso de comparación de dos grupos y una muestra pequeña (≤ 30) o prueba de chi, en caso de dos grupos con muestra ≥ 30 o más de dos grupos.

Variables de escala continua: Si distribuye normal, se comparan medidos. Si son dos medidos se usa la prueba de T student y si son tres o más medidos, la prueba de análisis de la varianza (ANOVA). Cuando no distribuye normal, se usa prueba de Mann Whitney comparación de dos medianas, Kursskall Wallies, comparación de tres o más medianas.

2° LA DETERMINACIÓN DE LA CORRELACIÓN ENTRE LA VARIABLE INDEPENDIENTE Y EL EFECTO, CON EL COEFICIENTE DE CORRELACIÓN.

La correlación es el grado de variación conjunta de dos variables numéricas o continuas. Se determina con un coeficiente de correlación que será de Pearson (r), en el caso que la variable distribuya normal, o de Spearman (r_{ho}) en el caso de que no distribuya normal.

Existen múltiples pruebas de comparación de grupos y de correlación para ser usados en variables con distribución normal o paramétricas como para aquellos que no tienen distribución normal o no paramétricas.

Para determinar el grado de asociación entre una variable independiente y una dependiente, se usa la regresión. Su forma más sencilla, es la regresión lineal simple, que es una técnica estadística que analiza la relación entre dos variables cuantitativas que distribuyen normal y donde se trata de verificar que existe una relación lineal.

Una asociación lineal, significa que a medida que la variable independientemente esta presente, también esta presente en la misma proporción variable dependiente.

MEDIDAS DE ASOCIACIÓN:

Son aquellas medidas que reportan la magnitud de asociación entre la variable explicadora X y la variable dependiente Y .

También dice la fuerza con la que se asocian ambas variables.

Las medidas, más frecuentemente utilizadas en los estudios, son:

- RIESGO RELATIVO (RR): Se puede estimar solo en estudios prospectivos.
- ODDS RATIO: Solo se calcula en estudios con respuesta dicotómica, sale de una regresión logística.
- HAZARD RATIO (HR): Se puede calcular en estudios prospectivos, es el resultado de la regresión de riesgos proporcionales de Cox.