



Nombre del alumno:

Rafael Torres Adorno

Nombre del profesor:

Lic. Rosario Gómez

Nombre del trabajo:

Métodos de muestreo e hipótesis

Materia:

Bioestadística

Grado:

Cuarto Cuatrimestre

Grupo:

Licenciatura Enfermería

Pichucalco, Chiapas a 13 de noviembre de 2020.

Introducción:

En la estadística la herramienta fundamental para poder realizar muestras de cada grupo se le conoce como estimaciones, el cual es lo que se tiene que comprobar los valores por métodos, el cual dará un valor aproximado de un parámetro que corresponde a una población, esto va a partir de los datos que se obtienen de las muestras recolectadas, claro que, para recolectarlo, existen herramientas que nos ayudan a eso.

La Estadística descriptiva y la teoría de la Probabilidad van a ser los pilares de un nuevo procedimiento (Estadística Inferencial) con los que se va a estudiar el comportamiento global de un fenómeno. La probabilidad y los modelos de distribución junto con las técnicas descriptivas, constituyen la base de una nueva forma de interpretar la información suministrada por una parcela de la realidad que interesa investigar.

Todos estos datos los podremos representar gráficamente, el cual nos servirán poder detectar en que parte de la población presenta cambios que ponen en situaciones a ellas, con estas podemos implementar herramientas o estrategias para dar posibles soluciones.

A continuación, se presentarán conceptos de estimaciones, y con la relación de la hipótesis, con el fin de conocer cómo se implementan.

Función de distribución empírica

Los tratamientos estadísticos se caracterizan por un ir y venir permanente entre los datos, que son colecciones de cifras medidas, y los modelos probabilistas que no tienen ninguna realidad física, pero proveen herramientas para describir la variabilidad de los datos. En esta manera de pensar, un primer paso consiste en asociar a la muestra una ley de probabilidad ficticia.

La distribución empírica asociada a una muestra es la ley de probabilidad sobre el conjunto de las modalidades, que afecta a cada observación con el peso $\frac{1}{n}$. La idea es la siguiente. La forma más simple sería sacar aleatoriamente nuevos valores a partir de los valores ya observados, respetando sus frecuencias. En otras palabras, se simularía la distribución empírica.

La media, la varianza y la desviación estándar pueden ser vistas como características probabilistas de la distribución empírica. La media de la muestra es la esperanza de su distribución empírica.

Para un carácter discreto, la moda de la distribución empírica es el valor que tiene la frecuencia más alta. Para un carácter continuo agrupado en clases de amplitudes iguales, hablamos de clase modal. Una distribución empírica se llama unimodal si la frecuencia maximal es significativamente mayor que las otras. Puede ser bimodal o multimodal en otros casos.

Estadísticas muestrales

Es lo que resulta de considerar todas las muestras posibles que pueden ser tomadas de una población. Su estudio permite calcular la probabilidad que se tiene, dada una sola muestra, de acercarse al parámetro de la población. Mediante la distribución muestral se puede estimar el error para un tamaño de muestra dado.

La distribución de muestreo de una estadística es la distribución de esa estadística, considerada como una variable aleatoria, cuando se deriva de una muestra aleatoria de tamaño n . Se puede considerar como la distribución de la estadística para todas las muestras posibles de la misma población de un tamaño

de muestra dado. La distribución del muestreo depende de la distribución subyacente de la población, la estadística que se considera, el procedimiento de muestreo empleado y el tamaño de muestra utilizado. A menudo existe un considerable interés en si la distribución muestral puede aproximarse mediante una distribución asintótica, que corresponde al caso límite ya que el número de muestras aleatorias de tamaño finito, tomadas de una población infinita y utilizadas para producir la distribución, tiende a infinito.

Estimación

Se conoce por estimación al conjunto de técnicas que permiten dar un valor aproximado de un parámetro de una población a partir de los datos proporcionados por una muestra. Por ejemplo, una estimación de la media de una determinada característica de una población de tamaño N podría ser la media de esa misma característica para una muestra de tamaño n .

La estimación se encuentra dividido en tres grandes bloques, cada uno de los cuales tiene distintos métodos que se usan en función de las características y propósitos del estudio: La estimación puntual, el cual se divide en tres momentos, Método de los momentos, Método de la máxima verosimilitud, Método de los mínimos cuadrados. Luego tenemos lo que son los intervalos, estimación por intervalos y estimación bayesiana.

Estimación puntual

El objetivo de la estimación puntual es aproximar el valor del parámetro desconocido (tiempo medio de ejecución de un algoritmo, altura media de las mujeres de una población, diferencia del resultado medio entre dos tratamientos médicos, proporción de gente que mejora con un tratamiento médico).

Para ello se utiliza la información de la muestra (x_1, x_2, \dots, x_n) , a través de un estimador.

Estimación por intervalos

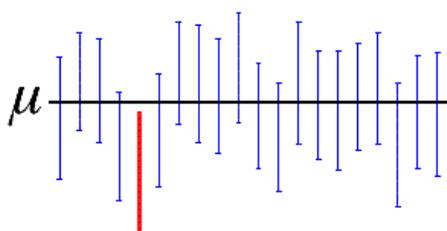
La estimación por intervalos consiste en establecer el intervalo de valores donde es más probable se encuentre el parámetro. La obtención del intervalo se basa en las siguientes consideraciones:

- a) Si conocemos la distribución muestral del estimador podemos obtener las probabilidades de ocurrencia de los estadísticos muestrales.
- b) Si conociéramos el valor del parámetro poblacional, podríamos establecer la probabilidad de que el estimador se halle dentro de los intervalos de la distribución muestral.
- c) El problema es que el parámetro poblacional es desconocido, y por ello el intervalo se establece alrededor del estimador. Si repetimos el muestreo un gran número de veces y definimos un intervalo alrededor de cada valor del estadístico muestral, el parámetro se sitúa dentro de cada intervalo en un porcentaje conocido de ocasiones. Este intervalo es denominado "intervalo de confianza".

Intervalos de confianza

Un intervalo de confianza es un rango de valores, derivado de los estadísticos de la muestra, que posiblemente incluya el valor de un parámetro de población desconocido. Debido a su naturaleza aleatoria, es poco probable que dos muestras de una población en particular produzcan intervalos de confianza idénticos. Sin embargo, si usted repitiera muchas veces su muestra, un determinado porcentaje de los intervalos de confianza resultantes incluiría el parámetro de población desconocido.

Para entender un poco de como se muestra gráficamente anexo un ejemplo con su explicación de ella misma.



En este caso, la línea negra horizontal representa el valor fijo de la media desconocida de la población, μ . Los intervalos de confianza azules verticales que se superponen a la línea horizontal contienen el valor de la media de la

población. El intervalo de confianza rojo que está completamente por debajo de la

línea horizontal no lo contiene. Un intervalo de confianza de 95% indica que 19 de 20 muestras (95%) de la misma población producirán intervalos de confianza que contendrán el parámetro de población.

Utilice el intervalo de confianza para evaluar la estimación del parámetro de población. Por ejemplo, un fabricante desea saber si la longitud media de los lápices que produce es diferente de la longitud objetivo. El fabricante toma una muestra aleatoria de lápices y determina que la longitud media de la muestra es 52 milímetros y el intervalo de confianza de 95% es (50,54). Por lo tanto, usted puede estar 95% seguro de que la longitud media de todos los lápices se encuentra entre 50 y 54 milímetros.

El intervalo de confianza se determina calculando una estimación de punto y luego determinando su margen de error.

Variabilidad de Parámetros

Este si no se conoce, puede obtenerse una aproximación en los datos o en un estudio piloto. También hay métodos para calcular el tamaño de la muestra que prescindan de este aspecto. Habitualmente se usa como medida de esta variabilidad la desviación típica poblacional y se denota σ .

Error de estimación

Un mismo estimador ofrece distintos valores para distintas muestras del mismo tamaño extraídas de la misma población. Por lo tanto, deberíamos tener una medida de la variabilidad del estimador respecto del parámetro que se trata de estimar. Esta variabilidad se mide en términos de la desviación estándar del estimador, la cual recibe el nombre de error estándar.

El error estándar de un estimador T de un parámetro es la desviación estándar del estimador.

Error de estimación es el valor absoluto de la diferencia entre una estimación particular y el valor del parámetro.

En realidad, por cada valor estimado del parámetro se tiene un error de estimación por lo general diferente. Sin embargo, es posible fijar un intervalo dentro del cual se encontrarán la mayoría de los valores de error de estimación para un estimador y parámetro dados.

Límite de confianza

Se le llama intervalo de confianza a un par o varios pares de números entre los cuales se estima que estará cierto valor desconocido con un determinado nivel de confianza. Formalmente, estos números determinan un intervalo, que se calcula a partir de datos de una muestra, y el valor desconocido es un parámetro poblacional. El nivel de confianza representa el porcentaje de intervalos que tomados de 100 muestras independientes distintas contienen en realidad el valor desconocido. En estas circunstancias, α es el llamado error aleatorio o nivel de significación, esto es, el número de intervalos sobre 100 que no contienen el valor.

Propiedades de los estimadores

Las propiedades o criterios para seleccionar un buen estimador son los siguientes:

A) Insesgadez: Diremos que un estimador θ^* de un parámetro θ es insesgado si su esperanza coincide con el verdadero valor del parámetro.

$$E[\theta^*] = \theta.$$

En el caso de que no coincidan, diremos que el estimador es sesgado.

B) Eficiencia: Dados dos estimadores θ_1^* y θ_2^* para un mismo parámetro θ , se dice que θ_1^* es más eficiente que θ_2^* si:

$$V[\theta_1^*] < V[\theta_2^*].$$

C) Suficiencia: Se dice que un estimador de un parámetro es suficiente cuando para su cálculo utiliza toda la información de la muestra.

D) Consistencia: Decimos que un estimador θ^* de un parámetro θ es consistente si la distribución del estimador tiende a concentrarse en un cierto punto cuando el tamaño de la muestra tiende a infinito.

Obtención de estimadores

El demostrar que un cierto estimador cumple estas propiedades puede ser complicado en determinadas ocasiones. Existen varios métodos que nos van a permitir obtener los estimadores puntuales. Los más importantes son:

- **MÉTODO DE LOS MOMENTOS:** se basa en que los momentos poblacionales y se estiman mediante los momentos muestrales. Suelen dar estimadores consistentes.
- **MÉTODO DE MÍNIMOS CUADRADOS:** consiste en obtener un estimador que hace mínima una determinada función.
- **MÉTODO DE MÁXIMA VEROSIMILITUD:** consiste en tomar como parámetro poblacional el valor de la muestra que sea más probable, es decir, que tenga mayor probabilidad. Se suelen obtener estimadores consistentes y eficientes. Es el más utilizado.

Estimación por intervalos de confianza

El intervalo de confianza está determinado por dos valores dentro de los cuales afirmamos que está el verdadero parámetro con cierta probabilidad. Son unos límites o margen de variabilidad que damos al valor estimado, para poder afirmar, bajo un criterio de probabilidad, que el verdadero valor no los rebasará. Es una expresión del tipo $[\theta_1, \theta_2]$ ó $\theta_1 \leq \theta \leq \theta_2$, donde θ es el parámetro a estimar. Este intervalo contiene al parámetro estimado con una determinada certeza o nivel de confianza.

Conclusión:

Las estimaciones nos ayudan a conocer como la población se encuentra ante diversas situaciones, el cual ciertas formulas son aplicadas para poder determinar cantidades de ellas.

Las estimaciones suelen dividirse en varios tipos, pero sea cual sea ese, tendrán distintos métodos, todo esto para cualquier tipo de estudio.

Algunos de los objetivos que estos temas tienen como propósito es el de calcular los parámetros de la distribución de medias o proporciones muestrales de tamaño n , extraídas de una población de media y varianza conocidas, el de estimar la media o la proporción de una población a partir de la media o proporción muestral, poder utilizar distintos tamaños muestrales para controlar la confianza y el error admitido. Poder contrastar los resultados obtenidos a partir de muestras, visualizar gráficamente, mediante las respectivas curvas normales, las estimaciones realizadas.

En la mayoría de las investigaciones resulta imposible estudiar a todos y cada uno de los individuos de la población ya sea por el coste que supondría, o por la imposibilidad de acceder a ello.

En definitiva, la idea es, a partir de una población se extrae una muestra por algunos de los métodos existentes, con la que se generan datos numéricos que se van a utilizar para generar estadísticos con los que realizar estimaciones o contrastes poblacionales.