UNIVERSIDAD DEL SURESTE

NOMBRE DE LA ALUMNA

Nelly Viridiana Díaz López

4TO CUATRIMESTRE GRUPO “A”

DOCENTE:

Juan Jesús Agustín Guzmán

MATERIA:

Bioestadística

TEMA:

Ensayo de la Unidad 3

TAPACHULA DE CORDOVA Y ORDOÑES, CHIAPAS.

A 23 DE OCTUBRE 2020

INTRODUCCIÓN

Este ensayo habla acerca de muestreo. Primero, El muestreo es la técnica de muestreo en la que todos los elementos forman el universo y que por lo tanto están descritos en el marco muestral, tienen idéntica probabilidad de ser seleccionados para la muestra.

Hablaremos de cómo vamos a justificar nuestro muestreo, de cómo vamos a se va a llevar a cabo cada función que se va a realiza.

Es algo corta mi introducción ya que es entre ensayo tratare de explicar más a fondo como el Muestreo nos puede ayudar en el ambiente de la carrera de Lic. En enfermería y no solo matemáticamente hablando.

- **Muestreo aleatorio simple (M.A.S.).**

Es la técnica de muestreo en la que todos los elementos forman el universo y que por lo tanto están descritos en el marco muestral, tienen idéntica probabilidad de ser seleccionados para la muestra. En la práctica, estos métodos pueden automatizarse mediante el uso de ordenadores, dependiendo de si los individuos del universo pueden ser seleccionados más de una vez en la muestra, hablaremos de M.A.S. con reposición o sin reposición.César Pérez López, nos dice en su libro "Muestreo Estadístico" (Pearson, 2005) como desarrolla de forma muy clara una comparación entre ambas técnicas. Tanto si lo miramos desde el punto de vista de qué técnica genera estimaciones más precisas como desde el punto de vista de qué técnica permite tener la misma precisión con menor tamaño de muestra, se puede concluir que el muestreo aleatorio simple sin reposición siempre es más eficiente. Para poder observar este resultado, partimos de la siguiente expresión para el tamaño de muestra en un M.A.S. sin reposición.

El muestreo aleatorio simple es un procedimiento de muestreo probabilístico que da a cada elemento de la población objetivo y a cada posible muestra de un tamaño determinado, la misma probabilidad de ser seleccionado. El muestreo aleatorio simple no es tan utilizado en investigaciones del consumidor, sobre todo porque es complicado obtener un marco de muestreo donde extraer al azar y no querrás darle a todas las unidades de la muestra una probabilidad igual de ser elegidas, ya que usualmente para hacer una investigación de este tipo se requiere a usuarios de tiendas o consumidores de ciertos productos o ciertas áreas específicas para ser las unidades de muestreo. No olvidemos que una parte muy importante del muestreo consiste en tener el tamaño de la muestra correcta, para no tener un error de muestreo, el cual debe ser el mínimo posible. Pasos para seleccionar una muestra aleatoria simple:

- Define la población objetivo
- Identifica un marco de muestreo actual de la población objetivo o desarrolla uno nuevo
- Evalúa el marco de muestreo para la falta de cobertura, cobertura excesiva, cobertura múltiple y la agrupación, y haz los ajustes que consideres necesarios.
- Asigna un número único a cada elemento de la trama.
- Determina el tamaño de la muestra.
- Selecciona al azar el número específico de elementos de la población.

- **Justificación del muestreo**.

Existen tres razones principales para extraer una muestra. Por lo general, lleva demasiado tiempo realizar un censo completo. En segundo lugar, es demasiado costoso hacer un censo completo. Tercero, es demasiado molesto e ineficiente obtener un conteo completo de la población objeto. Después que se han determinado las preguntas numéricas y categóricas más esenciales en la encuesta, el tamaño de muestra necesario se basará en la satisfacción de la pregunta con los requerimientos más rigurosos.

- **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. 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 1/n. 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 continúo 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. Para estudiar una distribución empírica, la primera etapa consiste en ordenar los datos en orden creciente, es decir escribir sus estadígrafos de orden. La función de distribución empírica (FED) o cdf empírica es una función de paso que salta por 1/N a la ocurrencia de cada observación. Por definición, la función FDE calcula la distribución acumulativa del número aleatorio subyacente. El FED estima la verdadera función de densidad acumulativa subyacente de los puntos en la muestra; Se garantiza virtualmente que converge con la distribución verdadera a medida que el tamaño de la muestra se hace lo suficientemente grande. Si se tiene una muestra aleatoria simple, es posible encontrar una distribución a partir de la muestra que proporcionará un cierto parecido a la distribución verdadera de la variable asociada con la población. Es lo que se denomina función de distribución empírica de la muestra. La principal propiedad de la función de distribución empírica de la muestra es su aproximación a la función de distribución poblacional cuando aumenta el tamaño muestral. Ello es conocido en estadística como el teorema de Glivenko-Cantelli o también como teorema central de la estadística.

- **Estadísticos muéstrales.**

Distribuciones: En estadística un estadístico (muestral) es una medida cuantitativa, derivada de un conjunto de datos de una muestra, con el objetivo de estimar o inferir características de una población o modelo estadístico. Más formalmente un estadístico es una función medible T que, dada una muestra estadística de valores, les asigna un número, que sirve para estimar determinado parámetro de la distribución de la que procede la muestra. Esto se denomina como realizar una estimación puntual. A partir de las muestras seleccionadas de una población pueden construirse variables aleatorias alternativas, de cuyo análisis se desprenden interesantes propiedades estadísticas. Las dos formas más comunes de estas variables corresponden a las distribuciones muéstrales de las medias y de las proporciones. Dada una población constituida por un número n de elementos, cuya media aritmética es m y donde la desviación típica viene dada s, pueden formarse n2 muestras con reemplazamiento distintas, formadas por dos elementos de la población. El estudio de determinadas características de una población se efectúa a través de diversas muestras que pueden extraerse de ella. El muestreo puede hacerse con o sin reposición, y la población de partida puede ser infinita o finita. Una población finita en la que se efectúa muestreo con reposición puede considerarse infinita teóricamente. También, a efectos prácticos, una población muy grande puede considerarse como infinita. En todo nuestro estudio vamos a limitarnos a una población de partida infinita o a muestreo con reposición. Consideremos todas las posibles muestras de tamaño n en una población. Para cada muestra podemos calcular un estadístico (media, desviación típica, proporción) que variará de una a otra. Así obtenemos una distribución del estadístico que se llama distribución muestral. Las dos medidas fundamentales de esta distribución son la media y la desviación típica, también denominada error típico. Hay que hacer notar que si el tamaño de la muestra es lo suficientemente grande las distribuciones muéstrales son normales y en esto se basarán todos los resultados que alcancemos.

- **Estimación.**

La estimación puntual consiste en atribuir un valor (la estimación) al parámetro poblacional. Si la muestra es representativa de la población, podemos esperar que los estadísticos calculados en las muestras tengan valores semejantes a los parámetros poblacionales, y la estimación consiste en asignar los valores de los 64 estadísticos muéstrales a los parámetros poblacionales. Los estadísticos con que obtenemos las estimaciones se denominan estimadores. En inferencia estadística se llama 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 divide en tres grandes

bloques, cada uno de los cuales tiene distintos métodos que se usan en función de las características y propósitos del estudio: Estimación puntual: Método de los momentos; Método de la máxima verosimilitud; Método de los mínimos cuadrados; Estimación por intervalos. Estimación bayesiana.
Un estimador es una regla que establece cómo calcular una estimación basada en las mediciones contenidas en una muestra estadística. Estimación puntual: Consiste en la estimación del valor del parámetro mediante un sólo valor, obtenido de una fórmula determinada. Estimación por intervalos: Consiste en la obtención de un intervalo dentro del cual estará el valor del parámetro estimado con una cierta probabilidad. En la estimación por intervalos se usan los siguientes conceptos: Intervalo de confianza. El intervalo de confianza es una expresión del tipo [θ_1, θ_2] ó $\theta_1 \leq \theta \leq \theta_2$, donde θ es el parámetro a estimar. Este intervalo contiene al parámetro estimado con un determinado nivel de confianza. A veces puede cambiar este intervalo cuando la muestra no garantiza un axioma o un equivalente circunstancial.
Variabilidad del Parámetro: Si no se conoce, puede obtenerse una aproximación en los datos aportados por la literatura científica o en un estudio piloto. También hay métodos para calcular el tamaño de la muestra que prescinde de este aspecto. Habitualmente se usa como medida de esta variabilidad la desviación típica poblacional y se denota σ. Error de la estimación: Es una medida de su precisión que se corresponde con la amplitud del intervalo de confianza. Cuanta más precisión se desee en la estimación de un parámetro, más estrecho deberá ser el intervalo de confianza y, si se quiere mantener o disminuir el error, más observaciones deberán incluirse en la muestra estudiada. En caso de no incluir nuevas observaciones para la muestra, más error se comete al aumentar la precisión. Se suele llamar E, según la fórmula $E = (\theta_2 - \theta_1)/2$. Límite de Confianza: Es la probabilidad de que el verdadero valor del parámetro estimado en la población se sitúe en el intervalo de confianza obtenido. El nivel de confianza se denota por $(1-\alpha)$, aunque habitualmente suele expresarse con un porcentaje $((1-\alpha)\cdot 100\%)$. Es habitual tomar como nivel de confianza un 95% o un 99%, que se corresponden con valores α de 0,05 y 0,01 respectivamente. Valor α: También llamado nivel de significación. Es la probabilidad (en tanto por uno) de fallar en nuestra estimación, esto es, la diferencia entre la certeza (1) y el nivel de confianza $(1-\alpha)$.

- **Propiedades de los estimadores.**

ESTIMADOR: Es un estadístico (es decir, es una función de la muestra) usado para estimar un parámetro desconocido de la población. Si se desea conocer el precio medio de un artículo (el parámetro desconocido) se recogerán observaciones del precio de dicho artículo en diversos establecimientos (la muestra) y la media aritmética de las observaciones puede utilizarse como estimador del precio medio. Para cada parámetro pueden existir varios estimadores diferentes. En general, escogeremos el estimador que posea mejores propiedades que los restantes, como insesgadez, eficiencia, convergencia y robustez (consistencia).

SESGO: Se denomina sesgo de un estimador a la diferencia entre la esperanza (o valor esperado) del estimador y el verdadero valor del parámetro a estimar. Es deseable que un estimador sea insesgado o centrado, es decir, que su sesgo sea nulo por ser su esperanza igual al parámetro que se desea estimar. Por ejemplo, si se desea estimar la media de una población, la media aritmética de la muestra es un estimador insesgado de la misma, ya que su esperanza (valor esperado) es igual a la media de la población. EFICIENCIA: Un estimador es más eficiente o preciso que otro, si la varianza del primero es menor que la del segundo.

CONVERGENCIA: Para estudiar las características de un estimador no solo basta con saber el sesgo y la varianza, sino que además es útil hacer un análisis de su comportamiento y estabilidad en el largo plazo, esto es, su comportamiento asintótico. Cuando hablamos de estabilidad en largo plazo, se viene a la mente el concepto de convergencia.

- **Obtención de estimadores.**

Método de los momentos: Se trata de un método de obtención de estimadores muy intuitivo. Consiste en igualar los momentos poblacionales (que sean función del o los parámetros a estimar) con los momentos muéstrales y despejar el parámetro a estimar. La principal ventaja de este método es su simplicidad. Sin embargo, aunque los estimadores así obtenidos son consistentes, en general, no son centrados ni eficientes. Además, en ciertos casos puede proporcionar estimaciones absurdas. Recordemos que la esperanza de una distribución uniforme comprendida entre dos valores a y b es el promedio de estos dos valores. Por tanto, para aplicar el método de los momentos para estimar b, igualaremos dicho promedio a la media aritmética.

Ventajas y desventajas del método de los momentos: El método de los momentos es bastante sencillo y brinda estimadores compatibles (debajo suposiciones muy débiles), aunque estos estimadores son a menudo sesgados. En algunos casos, cuándo estimamos parámetros de una familia conocida de distribuciones de probabilidad, este método es sustituido el método de máxima verosimilitud de Fisher, porque con máxima verosimilitud los estimadores tienen probabilidad más alta de ser cercanos a las cantidades que estimamos y son menos sesgadas. Aun así, en algunos casos las ecuaciones del método de máxima verosimilitud pueden ser intratables sin ayuda de ordenadores, mientras que el método de estimadores de los momentos puede ser más accesible y fácilmente calculado a mano. Las estimaciones por el método de los momentos pueden ser utilizadas como la primera aproximación a las soluciones de las ecuaciones de verosimilitud, y podemos encontrar sucesivas mejoras en las aproximaciones por el método de Newton-Raphson. De este modo el método de momentos puede ayudar a encontrar estimaciones del método de máxima verosimilitud.

- **Estimación por intervalos de confianza.**

La estadística inferencial es el proceso de uso de los resultados derivados de las muestras para obtener conclusiones acerca de las características de una población. La estadística inferencial nos permite estimar características desconocidas como la media de la población o la proporción de la población. Existen dos tipos de estimaciones: la estimación puntual y la estimación de intervalo. Una estimación puntual es el valor de un solo estadístico de muestra. Una estimación del intervalo de confianza es un rango de números, llamado intervalo, construido alrededor de la estimación puntual. El intervalo de confianza se construye de manera que la probabilidad del parámetro de la población se localice en algún lugar dentro del intervalo conocido. 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 muéstrales. 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". La estimación puntual trata el problema de estimar mediante un número el valor de una característica poblacional o parámetro θ desconocido (por ejemplo, la estimación del IPC de un determinado periodo).

BIBLIOGRAFÍA:

ANTOLOGÍA DE LA UDS. BIOESTADISTICA. UNIDAD III. MUESTREO ALEATORIO SIMPLE. PÁG: 54-74