



Nombre del alumno: Maricela Álvarez Ton.

Nombre del profesor: Omar Ulises Hernández Liévano

Licenciatura: Enfermería.

Materia: Bioestadística

Nombre del trabajo: Ensayo de la Unidad III



Introducción

Se considera muestreo la técnica todos los elementos que forman el universo y que por lo tanto están descritos en el marco muestral, tienen idéntica probabilidad de ser seleccionados para la muestra. Al proceso en el que se generan las muestras, una muestra es una parte (un conjunto) de la población y se desea que la muestra sea lo más representativa posible de la población de la que procede, o que cumpla con las características más básicas, que en ocasiones en la que no es posible o conveniente analizar a todos los elementos de la población, se selecciona una muestra, entendiendo como una parte representativa del total de la población. Por lo tanto, el muestreo es una herramienta de la investigación científica cuya función principal es determinar que parte de la población debe examinarse con la finalidad de hacer inferencias sobre la misma.

La muestra debe lograr una presentación adecuada de la población en la que se reproduzca de la mejor manera los rasgos esenciales de la misma, que son importantes para la investigación, para que esta sea representativa, y por lo tanto útil debe reflejar de manera exhaustiva o en su mayoría las similitudes y diferencias encontradas tienen idéntica probabilidad de ser seleccionados para la muestra, por el hecho de que a veces las poblaciones que se desean estudiar pueden ser relativamente extensas, y pueden resultar muy difíciles de estudiar.

Desarrollo

Muestreo aleatorio simple.

Es la técnica de muestreo en la que todos los elementos que conforman el universo esta técnica son equivalentes a hacer un sorteo entre los individuos del universo.

- Subconjunto de una muestra elegida de una población más grande.
- cada individuo tiene la misma probabilidad de ser elegido en cualquier etapa del proceso.
- Cada individuo se elige al azar y por pura casualidad.

César Pérez López, en su libro "Muestreo Estadístico" 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.

La siguiente expresión para el tamaño de muestra en un M.A.S. sin reposición.

- La fórmula relaciona el tamaño de muestra necesario cuando el universo es finito con el tamaño necesario cuando el universo es infinito.
- Beneficios del muestreo aleatorio simple.
- El desarrollo de la informática ha permitido que diseñar una muestra aleatoria simple sea extremadamente rápido y fiable.

Todas las personas participantes tienen la misma posibilidad de ser seleccionadas, y dado que conocemos el tamaño de la población (N) y el tamaño de la muestra (n), el cálculo es el siguiente:

Por ejemplo:

$$P=1 - \frac{N-1}{N} \cdot \frac{N-2}{N-1} \dots \frac{N-n}{N-(n-1)}$$

$$\text{Eliminando} = 1 - \frac{n}{N}$$

$$= \frac{n}{N}$$

$$= \frac{100}{1000}$$

$$= 10\%$$

Para no tener un error de muestreo, el cual debe ser el mínimo posible, que tiende ser los pasos para seleccionar una muestra aleatoria simple.

- Define la población objetivo. Quizá quieres leer: ¿Como encontrar a tu mercado 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 excesiva, cobertura excesiva, cobertura múltiple y agrupación, y haz los ajustes que consideres necesarios.
- Asigna un número único a cada elemento de la trama.
- Selecciona al azar el número específico de elementos de la población.

El método de aleatorio solo funciona bien con pequeñas poblaciones de la muestra, es poco práctico para su uso con poblaciones más grandes.

Los números que el investigador encuentra que no concuerdan con los números asignados a elementos de la población objetivo son ignorados; este proceso de la tabla es un proceso tedioso, consume tiempo y esto sin embargo no se recomienda para grandes poblaciones.

En vez de tomar un censo completo, los procedimientos de muestreo estadístico se han convertido en la herramienta preferida en la mayoría de las situaciones de investigación, en este caso existen tres razones principales para extraer una muestra. En segundo lugar, es demasiado costoso hacer un censo completo, y un tercer lugar, es demasiado molesto e ineficiente obtener un conteo completo de la población objeto.

Los tratamientos estadísticos se caracterizan por un ir y venir permanente entre los datos.

1. Son colecciones de cifras medidas.
2. Proveen herramientas para describir la variedad de los datos.
3. Primer dato consiste en asociar a la muestra una ley de probabilidad ficticia.
4. La variación y la desviación estándar pueden ser vistas como características probabilidades de la distribución empírica.

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 de la muestra es la esperanza de su distribución empírica.

- Una distribución empírica se llama unimodal si la frecuencia maximal es significativamente mayor que las otras, puede ser bimodal o multimodal en otros casos.

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.

la función de distribución empírica para datos continuos agrupados es una función de las frecuencias relativas acumuladas. Pero en este caso, más que usar una función escalera, lo que se hace es dibujar las frecuencias acumuladas en relación de los límites superiores de cada clase, para a continuación, unir los puntos con líneas rectas.

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, es lo que resulta de considerar todas las muestras posibles que pueden ser tomadas de una población, 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, las dos formas más comunes de estas variables corresponden a las distribuciones muestrales de las medias y de las proporciones, 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 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.

Las dos medidas fundamentales de esta distribución son la media y la desviación típica, también denominada error típico. Si el tamaño de la muestra es lo suficientemente grande las distribuciones muestrales son normales y en esto se basarán todos los resultados que alcancemos.

La estimación puntual consiste en atribuir un valor (la estimación) al parámetro poblacional. Si la muestra es representativa de la población, la estimación es un cálculo que se realiza a partir de la evaluación estadística. Dicho estudio suele efectuarse sobre una muestra y no sobre toda la población objetivo.

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 . Para llevar a cabo una estimación, entonces, es necesario primero contar con una serie de datos.

La estimación puede ser utilizada para diversos fines como los siguientes:

Calcular indicadores estadísticos como la media, la mediana y la moda. Esto, respecto a cualquier variable económica.

Calcular el valor de una variable determinada con base en otras. Por ejemplo, estimar la balanza comercial si contamos con los datos de las exportaciones e importaciones de un país.

Realizar proyecciones de una variable. Esto puede ser posible a partir de datos históricos. Por ejemplo, tomando el modelo de las expectativas adaptativas puede estimarse la inflación esperada por los consumidores en función a los incrementos de precios en el pasado.

Calcular el valor contable o valorar un activo, tomando en cuenta que tiene un periodo de vida útil.

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.

Estimación puntual. Sea X una variable poblacional con distribución $F\theta$, siendo θ desconocido. El problema de estimación puntual consiste en, seleccionada una muestra X_1, \dots, X_n , encontrar el estadístico $T(X_1, \dots, X_n)$ que mejor estime el parámetro θ . Una vez observada o realizada la muestra, con valores x_1, \dots, x_n , se obtiene la estimación puntual de θ , $T(x_1, \dots, x_n) = \hat{\theta}$.

Es un estadístico (es decir, es una función de la muestra) usado para estimar un parámetro desconocido de la población.

Por ejemplo, si se desea conocer el precio medio de un artículo (el parámetro desconocido) se recogerán observaciones del precio de dicho artículo en diversos establecimientos.

ESTIMADOR:

Es un estadístico (es decir, es una función de la muestra) usado para estimar un parámetro desconocido de la población. 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. 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. Luego, podemos construir sucesiones de estimadores y estudiar el fenómeno de la convergencia.

Comportamiento Asintótico: En el caso de las variables aleatorias, existen diversos tipos de convergencia, dentro de las cuales podemos distinguir:

- Convergencia en probabilidad (o débil).
- Convergencia casi segura (o fuerte).
- Convergencia en media cuadrática.
- Convergencia en distribución.

CONSISTENCIA:

También llamada robustez, se utilizan cuando no es posible emplear estimadores de mínima varianza, el requisito mínimo deseable para un estimador es que a medida que el tamaño de la muestra crece, el valor del estimador tiende a ser el valor del parámetro, propiedad que se denomina consistencia.

- Un estimador Q de φ es un estimador insesgado si $E(Q) = \varphi$ para todos los valores posibles de φ .
- Sesgo $(Q) = E(Q) - \varphi$.

Obtención de estimadores Se trata de un método de obtención de estimadores muy intuitivo. Básicamente, consiste en igualar los momentos poblacionales (que sean función del o los parámetros a estimar) con los momentos muestrales y despejar el parámetro a estimar.

Así, por ejemplo, la esperanza de una variable aleatoria se estimaría por la media muestral; la varianza, por la varianza muestral. La principal ventaja de este método es su simplicidad; Recordemos que la esperanza de una distribución uniforme comprendida entre dos valores a y b es el promedio de estos dos valores.

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.
- Cuando 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.

1.- la especificación de la variable aleatoria y su campo de variación.

2.- la especificación de su asignación de probabilidades, mediante la función de distribución. (Alternativamente mediante la cuantía o densidad, la F.C. o la F.G.M. (si existe).

- FUNCIONES DE DEFINICIÓN
- ESTRUCTURA FUNCIONAL
- FAMILIA DE DISTRIBUCIONES
- MODELO DE PROBABILIDAD
- PARÁMETROS DE LA DISTRIBUCIÓN
- INFERENCIA ESTADÍSTICA.
- MODELOS DISCRETOS

❖ DISTRIBUCIÓN DICOTÓMICA.

Si una variable aleatoria X sigue o tiene una distribución dicotómica de parámetro p se expresa como $X \sim D(p)$. Modeliza situaciones en las que:

- Se realiza una prueba
- Que sólo puede dar dos resultados posibles: A y \bar{A}
- La probabilidad del resultado A es $P(A) = p$ y la del resultado \bar{A} es $P(\bar{A}) = q = 1 - p$.
- En estas circunstancias la variable aleatoria X significa "nº de resultados A que se obtienen.

❖ DISTRIBUCIÓN BINOMIAL.

Si una variable aleatoria, X , sigue una distribución binomial de parámetros n y p se expresa como: $X \sim B(n, p)$.

❖ DISTRIBUCIÓN HIPERGEOMÉTRICA

Dada la siguiente situación:

Una población constituida por N individuos en total. De los cuales Np individuos son del tipo A, y Nq individuos son del tipo \tilde{A} . De forma que la proporción de individuos A que hay en la población es p , y la proporción de individuos de tipo \tilde{A} , es q ($p+q=1$).

DISTRIBUCIÓN DE POISSON

Formalmente: dada una variable aleatoria X con campo de variación. Situaciones que modeliza: Se observa la ocurrencia de hechos de cierto tipo durante un período de tiempo o a lo largo de un espacio, considerados unitarios. El tiempo (o el espacio) pueden considerarse homogéneos, respecto al tipo de hechos estudiados, al menos durante el período experimental; es decir, que no hay razones para suponer que en ciertos momentos los hechos sean más probables que otros.

MODELOS CONTINUOS

DISTRIBUCIÓN UNIFORME (DE V. CONTINUA)

Dada una variable aleatoria continua, X , definida en el intervalo $[a,b]$ de la recta real, diremos que X tiene una distribución uniforme en el intervalo $[a,b]$ cuando su función de densidad sea:

$$X \sim U([a, b]) \quad f(x) = \frac{1}{(b-a)} \text{ para } x \in [a, b].$$

La distribución binomial es una distribución de probabilidad discreta que cuenta el número de éxitos en una secuencia de n ensayos de Bernoulli independientes entre sí, con una probabilidad fija p de ocurrencia del éxito entre los ensayos, una probabilidad $q=1-p$. En la distribución binomial el anterior experimento se repite n veces, de forma independiente, y se trata de calcular la probabilidad de un determinado número de éxitos. Para $n = 1$, la binomial se convierte, de hecho, en una distribución de Bernoulli.

La probabilidad del éxito ha de ser constante. Esta se representa mediante la letra p . La probabilidad de que salga cara al lanzar una moneda es 0,5 y esta es constante dado que la moneda no cambia en cada experimento y las probabilidades de sacar cara es constante, Esta se representa mediante la letra $q = 1-p$. Es importante fijarse que, mediante esa ecuación, sabiendo p o sabiendo q , podemos obtener la que nos falte.

En teoría de la probabilidad y estadística, la distribución de probabilidad de una variable aleatoria es una función que asigna a cada suceso definido sobre la variable la probabilidad de que dicho suceso ocurra.

La distribución de probabilidad está completamente especificada por la función de distribución, cuyo valor en cada x real es la probabilidad de que la variable aleatoria sea menor o igual que x .

Una distribución discreta describe la probabilidad de ocurrencia de cada valor de una variable aleatoria discreta.

Una variable aleatoria discreta es una variable aleatoria que tiene valores contables, tales como una lista de enteros no negativos.

La distribución normal es una distribución con forma de campana donde las desviaciones estándar sucesivas con respecto a la media establecen valores de referencia para estimar el porcentaje de observaciones de los datos. Estos valores de referencia son la base de muchas pruebas de hipótesis, como las pruebas Z y t.

Esta curva se conoce como campana de Gauss y es el gráfico de una función gaussiana.

Algunos ejemplos de variables asociadas a fenómenos naturales que siguen el modelo de la normal son:

- ❖ Caracteres morfológicos de individuos como la estatura
- ❖ Caracteres fisiológicos como el efecto de un fármaco
- ❖ Caracteres sociológicos como el consumo de cierto producto por un mismo grupo de individuos
- ❖ Caracteres psicológicos como el cociente intelectual
- ❖ Nivel de ruido en telecomunicaciones
- ❖ Errores cometidos al medir ciertas magnitudes

Inferencia estadística, un contraste de hipótesis (también denominado test de hipótesis o prueba de significación) es un procedimiento para juzgar si una propiedad que se supone en una población estadística es compatible con lo observado en una muestra de dicha población.

- ❖ Especifiquen un valor concreto o un intervalo para los parámetros del modelo.
- ❖ Determinen el tipo de distribución de probabilidad que ha generado los datos.

El contraste de hipótesis no establece la verdad de la hipótesis, sino un criterio que nos permite decidir si una hipótesis se acepta o se rechaza, o el determinar si las muestras observadas difieren significativamente de los resultados esperados. En este proceso podemos incurrir en dos tipos de errores según sea la situación real y la decisión que tomemos.

Una hipótesis estadística es una afirmación respecto a alguna característica de una población. Contrastar una hipótesis es comparar las predicciones con la realidad que observamos. Si dentro del margen de error que nos permitimos admitir, hay coincidencia, aceptaremos la hipótesis y en caso contrario la rechazaremos.

Construcción de Test de hipótesis.

- ❖ Seis pasos básicos para configurar y realizar correctamente una prueba de hipótesis.

- ❖ Especificar las hipótesis.
- ❖ Elegir un nivel de significancia (también denominado alfa o α).
- ❖ El gerente selecciona un nivel de significancia de 0.05, que es el nivel de significancia más utilizada.
- ❖ Determinar la potencia y el tamaño de la muestra para la prueba.

Contraste de hipótesis paramétricas.

El problema del contraste de hipótesis consiste básicamente en comprobar cotejar, decidir, en definitiva, sobre la veracidad de una hipótesis prefijada previamente como supuestamente cierta.

Una prueba de chi-cuadrada es una prueba de hipótesis que compara la distribución observada de los datos con una distribución esperada de los datos.

Existen varios tipos de pruebas de chi-cuadrada:

- ❖ Prueba de bondad de ajuste de chi-cuadrada
- ❖ Utilice este análisis para probar qué tan bien una muestra de datos categóricos se ajusta a una distribución teórica.

La bondad de ajuste de un modelo estadístico describe lo bien que se ajusta un conjunto de observaciones. Las medidas de bondad en general resumen la discrepancia entre los valores observados y los valores esperados en el modelo de estudio. Tales medidas se pueden emplear en el contraste de hipótesis, e.g.

Para calcular si una distribución dada se ajusta a un conjunto de datos, se pueden utilizar las siguientes pruebas:

- Prueba de Kolmogórov-Smirnov
- Criterio de Cramér-von Mises
- Prueba de Anderson-Darling
- Test de Shapiro–Wilk
- Prueba de ji cuadrada
- Criterio de Información de Akaike

Test de heterogeneidad.

Cuando hablamos de heterogeneidad podemos distinguir dos aspectos: por un lado, el relativo a las diferencias existentes entre los estudios en cuanto a características de los pacientes incluidos, la metodología utilizada, el tiempo de seguimiento, las dosis empleadas, la localización geográfica.

Tablas de Contingencia.

En estadística las tablas de contingencia se emplean para registrar y analizar la asociación entre dos o más variables, habitualmente de naturaleza cualitativa (nominales u ordinales).

Una tabla de contingencia es una tabla que cuenta las observaciones por múltiples variables categóricas. Las filas y columnas de las tablas corresponden a estas variables categóricas.

Las tablas de contingencia también pueden revelar asociaciones entre las dos variables.

Las tablas de contingencia más simples son tablas de dos factores que cuentan las respuestas según dos variables. Usted puede categorizar las observaciones según tres o más variables al "cruzarlas".

Una tabla de contingencia es una de las formas más comunes de resumir datos categóricos. En general, el interés se centra en estudiar si existe alguna asociación entre una variable denominada fila y otra variable denominada columna y se calcula la intensidad de dicha asociación.

De manera formal, se consideran X e Y dos variables categóricas con I y J categorías respectivamente. Una observación puede venir clasificada en una de las posibles $I \times J$ categorías que existen.

Cuando las casillas de la tabla contienen las frecuencias observadas, la tabla se denomina tabla de contingencia.

Conclusión.

En este capítulo describimos las principales conclusiones del trabajo. En la primera parte se realiza un análisis del grado en que se alcanzaron los objetivos propuestos en el trabajo y posteriormente, se proponen algunas modificaciones y sugerencias a la propuesta de actividades que surgen de las observaciones realizadas en la implementación, esto con el objetivo de mejorar el proceso de instrucción matemática que ésta promueve. Como se mencionó en su momento, el objetivo general de este trabajo consiste en diseñar una propuesta de actividades didácticas que permita promover un acercamiento intuitivo a algunos tipos de muestreos. El propósito general de la propuesta de actividades didácticas parte de que el estudiante construya su significado del muestreo a partir de un sistema de prácticas que le permitan identificar cuando un muestreo es aleatorio y cuando no lo es, además que identifique las diferencias entre los principales tipos de muestreo aleatorio, así como valorar la pertinencia de aplicar cierto tipo de muestreo en una situación específica.

El primer objetivo específico fue:

- a) Identificar si una propuesta de muestreo es aleatoria o no.

Con base en el análisis de la implementación de la propuesta de actividades, se puede constatar que la mayoría de los estudiantes identificaron apropiadamente los muestreos realizados al clasificarlos como aleatorios o no aleatorios, lo cual se percibe en su participación en las actividades de desarrollo y de cierre. En la actividad de cierre, que estaba más orientada a la evaluación, los estudiantes identificaron adecuadamente el tipo de muestreo aleatorio que se realizó en los casos correspondientes.

El segundo objetivo específico fue:

- b) Realizar muestreos aleatorios y no aleatorios para estimar la media poblacional.

A través de los análisis realizados, pudimos constatar que este objetivo se alcanzó de forma satisfactoria. Lo que permitió valorarlos como más pertinentes para estimar este parámetro.

El tercer objetivo específico fue:

- c) Determinar la pertinencia de utilizar cierto tipo de muestreo aleatorio en una situación específica.

Se considera que las dificultades que se presentaron respecto al manejo de Excel para la realización del muestreo aleatorio estratificado y por conglomerados. A pesar de las dificultades mencionadas, consideramos que la propuesta incide en buena medida en la construcción de un significado personal del muestreo, donde tras la realización del sistema de prácticas que promueve dicha propuesta, el estudiante

puede identificar, proponer y valorar la pertinencia de los muestreos aleatorios y no aleatorios al momento de estimar algún parámetro poblacional. Siendo la presente propuesta un primer acercamiento para el logro de los objetivos planteado, se considera pertinente proponer las siguientes modificaciones para enriquecer las prácticas que deberán realizar los estudiantes con el propósito de incidir en dichos objetivos.

Bibliografía.

título del libro: bioestadística wayne w. danielimusa sa de cv 2011.

título del libro: bioestadística alfredo de jesus celis de la rosa el manual moderno 2008.

<http://tesis.uson.mx/digital/tesis/docs/22068/Capitulo5>.