
1. Introducción

La inferencia estadística permite generalizar lo hallado en una muestra a toda la población. Para realizar este proceso contamos con dos opciones: Estimación y las Pruebas de hipótesis. El fundamento de estos procesos está relacionado con varios conceptos que se describen brevemente a continuación:

1.1 Población:

En Estadística, se identifica el término población con el dominio de la variable aleatoria X , asociada a los objetos o individuos sobre los cuales se desarrolla un experimento y cuyo valor ocurre al azar. El estudio del conjunto de todas las mediciones de interés para un investigador se llama CENSO (Mendenhall (2008)). Como ejemplo podemos mencionar la población de habitantes de Colombia que se estudia a través del Censo de población que se debe realizar cada diez años.

1.2 Muestra:

Un subconjunto extraído de los elementos que conforman la población se denomina MUESTRA. Una definición técnica de muestra está dada por: repetición n veces, en idénticas condiciones de la experiencia aleatoria, se puede obtener n valores independientes de una variable aleatoria X_1, X_2, \dots, X_n a la que se le denomina muestra de la variable X .

Una muestra de tamaño n , extraída de esta población en la que se observa el color de las correspondientes bolas, deberá ser una secuencia de unos y ceros, en total n . Para garantizar que todas las repeticiones se realicen en iguales condiciones se puede optar por regresar la bola extraída después de ser observada o también considerando un gran tamaño de bolas.

1.3 Parámetro:

Es una caracterización numérica de la distribución de probabilidad de una variable aleatoria. Como ejemplo de parámetros tenemos a μ, σ^2 que determinan la función de probabilidad de una variable con distribución normal. Si suponemos que los valores correspondientes a estos parámetros son 100 y 25 respectivamente, entonces la función de distribución de probabilidad quedará determinada por :

$$f(x) = \frac{1}{\sqrt{50\pi}} \exp \left(-\frac{1}{50} (x - 100)^2 \right)$$

note que $\theta_1 = 100$ es parámetro de centramiento y $\theta_2 = 25$ es parámetro de variación.

1.4 Estimador:

Es una función de los valores obtenidos en una muestra aleatoria que da como resultado un valor que corresponde a una aproximación del parámetro objeto de estudio. Generalmente se representa como $\hat{\theta}$. Como algunos ejemplos podemos citar:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$$

$$\widehat{\sigma^2} = \frac{1}{n-1} \sum_{i=1}^n (x - \bar{x})^2 = s^2$$

1.5 Estimación:

Es la evaluación o generación del estimador para una muestra determinada.

Como ejemplo podemos utilizar uno de los estimadores que estudiaremos : $\hat{\mu} = \bar{x}$, para una muestra dada

630, 650, 710, 750, 790, 820, 860 y 910

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{8}(630 + 650 + \dots + 910) = 765$$

El valor del estimador de μ para esta muestra es

$$\hat{\mu} = \bar{x} = 765$$

2. TÉCNICAS DE MUESTREO

2.1 Importancia del Muestreo.

El papel principal de la estadística está relacionado con el análisis de información que permita tomar decisiones con respecto a una población. Es posible que se pueda estudiar la totalidad de elementos que determinan la población, en este caso estaremos realizando un censo. Sin embargo no siempre podemos estudiar la totalidad de la población por lo cual debemos seleccionar una pequeña parte de ella. A este subconjunto lo podemos llamar muestra.

La siguiente tabla muestra las condiciones favorables o no relacionadas con la decisión de hacer censo o hacer muestreo.

Condiciones favorables en el uso del censo o muestra		
	MUESTRA	CENSO
Presupuesto	pequeño	grande
Tiempo disponible	poco	mucho
Tamaño e la población	grande	pequeña
Varianza de la característica	grande	pequeña
Costo de los errores de muestreo	bajo	alto
Costo de los errores que no son de muestreo	alto	bajo
Naturaleza de la prueba	destructiva	no destructiva
Atención a casos individuales	si	no

Tomada de Malhotra(2004)

Un diseño de muestreo es una estrategia para obtener una muestra, genera algunas preguntas asociadas al proceso de muestreo: ¿cómo seleccionar de una manera óptima la muestra?, ¿Qué característica medir en las unidades observadas? y ¿cómo estimar las características poblacionales a partir de la información muestral?. El proceso de obtención de las muestras por su parte, requiere la definición de algunos aspectos como, ¿cuál es el tamaño óptimo para el cumplimiento de los objetivos?, o ¿mediante qué procedimiento de aleatorización realizar el proceso de selección?, ¿qué tipo de método observacional utilizar y qué medidas tomar?. En el muestreo, uno tiene la oportunidad de seleccionar deliberadamente la muestra, para evitar inducir por ejemplo, a seleccionar por conveniencia algún individuo en particular, situación que pudiera conducir a sesgos en los resultados finales de la investigación.

En resumen, podemos decir que el muestreo consiste en la selección, mediante procedimientos preestablecidos que nos aseguren aleatoriedad y representatividad, de una parte de una colección finita de unidades, seguida de algunas conclusiones respecto de la población, basándonos en la parte de ella que hemos observado. Este proceso se basa en principios relacionados con la inferencia estadística, lo que implica fuentes de incertidumbre (producto de la aleatoriedad), como por ejemplo,

El método de selección de las unidades, el de medición de las unidades y el conocimiento de los procesos que generan los verdaderos valores de las unidades medidas.

Cada una de estas fuentes de variación, contribuye a la estructura estocástica del proceso inferencial. La primera se refiere a una estructura estocástica introducida cuando un mecanismo de aleatorización que determina qué parte de la colección de valores es seleccionado.

Una vez que la unidad ha sido seleccionada, debe ser utilizada para obtener una medida de dicha variable, el ejemplo (b) se refiere a aquellas situaciones en las que dichas medidas son imperfectas y contienen una fuente de variación propia de la medición. Este tipo de errores se puede producir por distintos motivos. Entre éstos están, los errores de los propios instrumentos de medición, los cuales tienen una precisión y error de medición propios. Dentro de este mismo proceso están las condiciones de medición, las que cuando no son favorables introducen error e incertidumbre, la que se suma a la variabilidad propia del proceso muestral. Una tercera fuente de incertidumbre en la medición la constituye la habilidad del muestreador para operar los instrumentos de medición.

Cuando hablamos de incertidumbre del proceso muestral estamos, por un lado, resumiendo todas las fuentes de error antes mencionadas y, por otro, debemos considerar la posibilidad que en el proceso de selección aleatoria de la muestra ésta incluya, por ejemplo, sólo las unidades mayores (o las menores) de la población, haciendo que la muestra no sea realmente representativa.

En el tercer caso, supongamos que la medición se hace en dos oportunidades, separadas por un año, luego, la medida de este año depende de la medida del año anterior, entonces, basta conocer la medida del año recién pasado más un modelo estocástico que relacione las medidas de este año con las medidas anteriores.

2.2 Ideas Básicas de Muestreo y Estimación

Cuando hablamos de obtener conclusiones respecto de una población particular, o más específicamente cuando intentamos caracterizar a una población, desde una perspectiva estadística, nos estamos refiriendo a algunas características distribucionales de la población. Específicamente, nos referimos a su distribución de probabilidades $F(\theta)$ y más específicamente aún, a los parámetros θ que la caracterizan.

Para el análisis que haremos a continuación, y en general para el resto del curso, asumimos dos supuestos básicos,

Los parámetros de la distribución $F(\theta)$ son desconocidos, ya que de no ser así no tendríamos un problema estadístico que resolver. y si las poblaciones son censables no tendríamos un problema inferencial, sino descriptivo.

De lo anterior concluimos que la única forma de obtener información sobre la distribución de probabilidades de la población es por medio de una muestra. Respecto de ésta, al estar la población caracterizada por su distribución de probabilidades, tenemos que cada elemento de la población, y por tanto de la muestra, contiene información acerca de dicha distribución de probabilidades. Luego, si deseamos conocer la población, apartir de una muestra de ella, y a partir de la información contenida allí, podemos, inferencialmente, y a partir de la construcción de estimadores $\hat{\theta}$ del parámetro θ de la distribución, obtener un estimador $\widehat{F(\theta)} = F(\hat{\theta})$. Esto significa que la inferencia en cuestión será relativa a un conjunto de parámetros poblacionales. De aquí que se habla también de inferencia paramétrica.

Como un ejemplo, cada hora, una máquina pinta un tanque, y algunos tanques son pintados de manera insuficiente. Un trabajador asignado 3 días a la semana, en 2 horas del día, debe verificar la pintura en la totalidad del área del tanque. La verificación de la pintura de cada tanque podemos pensarlo como un ensayo Bernoulli, donde éxito corresponde a una pintura insuficiente (con probabilidad θ) y fracaso pintura suficiente (con probabilidad $1 - \theta$). Si consideramos que estos ensayos son independientes, para una sucesión de n horas de revisión, observamos Y_1, Y_2, \dots, Y_n variables aleatorias Bernoulli independientes, cada una de parámetro θ , donde $Y_i = 1$ si el pintado es insuficiente en una hora y $Y_i = 0$ si no. Nos preguntamos ahora, ¿cómo esta sucesión de ceros y unos podemos utilizarla para obtener información acerca de θ , la probabilidad de pintura insuficiente cada hora?. En este caso $\sum y_i$ corresponde al número de días, del total de n , en el que la pintura fue insuficiente por lo que parece intuitivamente claro que $\bar{y} = \frac{1}{n} \sum y_i$ sea

una buena aproximación para θ .

En este ejemplo hemos observado valores de variables aleatorias independientes Y_1, Y_2, \dots, Y_n , donde cada una de ellas tiene la misma distribución de probabilidades. En estos casos, hablamos de muestra aleatoria de tamaño n . Formalmente, si Y_1, Y_2, \dots, Y_n son variables aleatorias independientes y están idénticamente distribuidas, cada una con la misma distribución de alguna variable aleatoria Y , entonces llamamos a Y_1, Y_2, \dots, Y_n una muestra aleatoria (m.a.) de la variable aleatoria Y . Si Y_1, Y_2, \dots, Y_n es una muestra aleatoria de una variable aleatoria Y , entonces se acostumbra a llamar a Y variable aleatoria poblacional o sencillamente población.

Cuando Y_1, Y_2, \dots, Y_n es una muestra aleatoria de una población Y , y conocemos la distribución de probabilidades de Y , entonces conocemos también la distribución conjunta de Y_1, Y_2, \dots, Y_n y podemos evaluar inmediatamente la distribución de cada Y_i . Generalmente, por supuesto, uno o más aspectos de la distribución de probabilidades para la población serán desconocidos para nosotros, por ejemplo él o los parámetros que caracterizan a la distribución. Incluso podemos desconocer su función de densidad o función de probabilidad, y por lo tanto, debemos sólo suponer que Y_1, Y_2, \dots, Y_n son independientes. Nuestro interrogante será saber si tienen alguna distribución específica, como una distribución normal, exponencial, etc. y buscar formas de verificarlo.

Nuestro propósito en la inferencia, es utilizar los elementos de la muestra para determinar todos los aspectos desconocidos de nuestro interés, en la mejor forma posible, al considerar la distribución de probabilidades de la población.

2.2.1 Diseños Muestrales

Un diseño muestral es una estrategia de selección de unidades muestrales, mediante un proceso de aleatorización definido previamente (plan de muestreo). De acuerdo con esta definición, son tres los elementos esenciales de un diseño muestral:

2.2.2 Unidad muestral.

Constituye la unidad básica a partir de la cual se obtiene la información, pudiendo por tanto, ser éstas personas, o individuos de cualquier tipo si, por ejemplo nuestro interés es estimar la talla promedio; casas, número promedio de habitaciones, o el consumo de energía; comunidades, si lo que nos interesa es el número promedio de especies por comunidad, áreas o cuadrículas si deseamos estimar densidades medias o biomasa total, etc.

2.2.3 Proceso de aleatorización

Aun cuando la unidad muestral esté claramente definida, existen algunos elementos que pueden afectar el proceso de muestreo, alterando la calidad de la información. En este caso no es la

definición, sino la selección de la unidad muestral, la que puede estar sesgada por la naturaleza de la medición, así por ejemplo, si deseamos encuestar a personas para averiguar su tendencia política o su nivel socio económico, pudiéramos sentirnos tentados a seleccionar sus nombres desde una guía telefónica, lo que dejaría fuera de la encuesta a todas aquellas personas que no tienen teléfono, constituyéndose ello en una fuente de sesgo. En otros casos, la naturaleza del muestreo tiene que ver con el entrenamiento de los muestreadores, es el caso, por ejemplo, de las muestras para identificar fallas en un equipo, en que la experiencia del muestreador juega un papel importante en la identificación de los mismos.

Lo anterior significa que, en términos prácticos, la aleatoriedad debe garantizar que cada individuo de la población debe tener la misma posibilidad de ser seleccionado como representante de la población (aleatoriedad). Adicionalmente, exigimos que la elección de un individuo, no esté condicionada a la selección de otro (independencia). Estas dos condiciones, en apariencia sencillas, pueden provocar algunas dificultades.

Supongamos por ejemplo, que estamos investigando el efecto de un cierto tipo de metal pesado como contaminante, cuyo efecto no será evidente sino en al menos 10 años. ¿Cómo seleccionamos la muestra correspondiente para medir su efecto?. Si deseamos medir el efecto de este mismo contaminante sobre la vida de los individuos que afecta, ¿de qué forma deben ser seleccionadas?, sabiendo que algunos de los individuos afectados pueden vivir 25 o más años. ¿Cuánto tiempo deberemos esperar para dar por concluido el proceso de muestreo?. Supongamos, por otra parte, que deseamos conocer la opinión política de un determinado grupo de individuos, ¿bajo qué circunstancias debemos formular las preguntas, para asegurar que los entrevistados respondan libremente a las preguntas, sin estar influenciados, por ejemplo, por la presencia de sus jefes o superiores que pudieran pensar diferente?. Estas y otras interrogantes nos sugieren la necesidad de formular planes y estrategias de muestreo adecuadas para cada una de las circunstancias, de forma de permitir, precisamente, la aleatoriedad y la independencia de las mediciones. Este es el objetivo del estudio de las técnicas de muestreo.

2.2.4 Tamaño muestral

Una vez que las unidades muestrales han sido definidas y se ha acordado un proceso de aleatorización, debemos preguntarnos, cuántas unidades debemos seleccionar para tener una buena representación en la muestra.

Dado que cualquiera sea el proceso de selección de una muestra existe un costo asociado, el que en la mayoría de los casos constituye una exigencia ineludible, nuestro objetivo principal, será obtener el máximo de información con el menor tamaño de muestra que nos sea posible.

El tamaño de la muestra n es función de tres elementos que son: la varianza, la confiabilidad y

el error de muestreo. Los dos últimos a criterio del investigador.

$$n = \frac{z_{\alpha/2}^2 \sigma^2}{e^2}$$

Donde:

Varianza σ^2 : Caracteriza la variable a estimar. Entre mayor sea su valor, mayor deberá ser el tamaño de la muestra. En ocasiones es necesario realizar una prueba piloto que nos permita tener un valor para calcular n . Otra alternativa puede ser el conocimiento empírico de expertos sobre el rango de la variable. Con este valor podemos estimar una aproximación de la desviación estándar como: $\sigma = \text{rango}/4$.

Confiabilidad: Este concepto está relacionado con el grado de veracidad que tienen los resultados obtenidos. Si el estudio es repetido muchas veces, ¿cuántas de estas coinciden con los resultados obtenidos?. Su valor está relacionado con el percentil de la distribución normal estándar, por ejemplo: una confiabilidad del 95 % está relacionada con un valor de $z = 1,96$.

Error de muestreo: corresponde a la diferencia entre el valor de la característica en la población (parámetro) y el valor obtenido con la información de la muestra (estimador). Equivale al error que estamos en capacidad de tolerar en las unidades de la variable.

Ejemplo de cálculo de tamaño de la muestra- Caso de una media

Suponga que deseamos determinar el tamaño de la muestra para estimar la media de la nota del curso de estadística, con una confianza del 95 % y un error de muestreo de 0.2 puntos. Un estudio realizado el semestre pasado arrojó una varianza de 1.6.

La información necesaria para el cálculo del tamaño de muestra requerido es: El percentil 97.5 de la distribución normal $z = 1,96$ (Entre -1.96 y 1.96 se encuentra el 95 % de los datos), la varianza $\sigma^2 = 0,4$ y el error de muestreo 0.2

$$n = \frac{1,96^2 \times 1,6}{0,2^2} = 153,6 \approx 154$$

En caso de que $n/N > 0,05$ debemos realizar un ajuste por corrección por población finita:

$$n = \frac{n_0 \times N}{n_0 + N - 1}$$

Supongamos que $N=200$ estudiantes y por tanto $154/200 > 0,05$ El tamaño corregido será:

$$n = \frac{154 \times 200}{154 + 200 - 1} = 87,2 \approx 88$$

Ejemplo de cálculo de tamaño de la muestra- Caso de una proporción

Supongamos ahora que se desea calcular el tamaño de la muestra para realizar una estimación de la proporción de estudiantes que reprueban el curso de matemáticas fundamentales y que deseamos una confianza del 98 % y un error de muestreo de 0.1. En este caso la varianza corresponde a pq donde $q = 1 - p$ y por tanto tendrá su valor máximo cuando $p = 0,50$, es decir que si utilizamos como varianza $0,5 \times (1 - 0,5) = 0,25$ obtendremos el valor más grande de tamaño de muestra posible.

$$n = \frac{2,33^2 \times 0,25}{0,1^2} = 135,7 \approx 136$$

En caso de poder realizar una prueba piloto (observación de una muestra, digamos 30 unidades) y estimar en este caso su varianza, el resultado del tamaño de muestra será inferior al anteriormente calculado.

3. Tipos de muestreo

En general pueden dividirse en dos grandes grupos: métodos de muestreo probabilísticos y métodos de muestreo no probabilísticos.

Los métodos de muestreo probabilísticos son aquellos que se basan en el principio de equiprobabilidad. Es decir, aquellos en los que todos los individuos tienen la misma probabilidad de ser elegidos para formar parte de una muestra y, consiguientemente, todas las posibles muestras de tamaño n tienen la misma probabilidad de ser elegidas. Sólo estos métodos de muestreo probabilísticos nos aseguran la representatividad de la muestra extraída y son, por tanto, los más recomendables. Dentro de los métodos de muestreo probabilísticos encontramos los siguientes tipos: De acuerdo con las consideraciones anteriores, algunas de las estrategias de muestreo más conocidas son: muestreo aleatorio simple, muestreo aleatorio sistemático, muestreo aleatorio estratificado, muestreo aleatorio por conglomerados.

3.1 Muestreos probabilísticos

3.1.1. Muestreo Aleatorio simple (m.a.s)

Una muestra aleatoria simple tomada de una población finita, da a cada subconjunto de la población (muestra) de tamaño específico la misma probabilidad de ser seleccionada.

El muestreo aleatorio simple es un procedimiento de selección de un conjunto de individuos de modo que toda posible muestra de n unidades tiene la misma probabilidad de ser seleccionada de los N individuos asociados a la población. La muestra se selecciona en n etapas, en cada una de las cuales, cada elemento que no ha sido aún seleccionado, tiene la misma probabilidad de ser seleccionado. Si el número de individuos de donde seleccionamos los n individuos para determinar la muestra tiene un tamaño conocido, podemos asignar un número a cada una de las

N unidades y luego mediante la generación de n números aleatorios entre 1 y N , seleccionamos los individuos que determinarán la muestra de tamaño n correspondiente.

El muestreo puede realizarse sin reemplazo o con reemplazo, o dicho de otro modo, la muestra puede elegirse de dos maneras: sin reposición de la unidad extraída o con reposición de ella. El muestreo con reemplazo es usado cuando es aceptable tener la misma unidad dos veces en la muestra.

Ejemplo

De la base del caso 1 con 200 observaciones deseamos seleccionar una muestra aleatorio de tamaño $n=50$. La siguiente es la instrucción en R que selecciona las observaciones :

```
# Importa los datos en formato csv
Base1=read.csv("~/Talleres/Base1.csv")
# visualiza los datos
viewData(Base1)
# extrae la muestra
muestra1=Base1[sample(1:nrow(Base1),50,replace=F),]
```

3.1.2. Muestreo Estratificado

Una muestra estratificada se obtiene formando estratos, de acuerdo a alguna característica de interés en la población (sexo, edad, ingresos etc.), y de cada estrato se selecciona una muestra aleatoria simple.

Supongamos, que la población finita se subdivide en L subpoblaciones o estratos, de modo que la muestra esté constituida por elementos de cada uno de ellos. El muestreo estratificado considera la selección, desde cada uno de los estratos, una muestra aleatoria simple.

La subdivisión de la población en estratos tiene como objetivo reducir la variabilidad total asociada al proceso de muestreo. Por lo tanto es necesario incluir en cada estrato, individuos o unidades muestrales cuya medida de interés tengan variabilidad pequeña respecto a la media. El objetivo general es entonces lograr varianza mínima dentro del estrato y varianza máxima entre estratos. En caso de no lograr esta estratificación, en función de las varianzas, lo más probable es que el proceso de muestreo sea equivalente al que se pudiera haber realizado por muestreo aleatorio simple.

La ventaja principal que puede conseguirse estratificando, es aumentar la precisión de las estimaciones al agrupar elementos con características comunes y con ello disminuir los tamaños muestrales totales y la amplitud de los intervalos de confianza. Por este tipo de muestreo, la amplitud del intervalo que estima al parámetro es menor que el dado por el m.a.s. para un mismo tamaño de muestra.

Por otra parte, la estratificación procura muestras más representativas. Exagerando, si se pudiera conseguir que cada uno de los L estratos estuviese constituido por elementos idénticos, bastaría tomar L elementos, uno por estrato, y la representatividad sería perfecta. Además, puede lograrse un mejor aprovechamiento de la organización administrativa destinada a la selección de las muestras, al centrar la actividad de muestreo en áreas homogéneas.

El número de estratos depende del criterio de clasificación, es un problema intrínsecamente ligado a la investigación, por ejemplo, si se está trabajando con características de individuos, los estratos pueden elegirse en la siguiente forma:

Estrato 1 ... niños (menores de 14 años)
Estrato 2 ... jóvenes (15 - 25 años)
Estrato 3 ... adultos (26 años o mas)

Otros criterios de clasificación pueden obedecer también, por ejemplo, a posiciones geográficas, en otros casos, cada estrato puede ser un sector de la ciudad de la provincia o de la región.

Otro elemento interesante de considerar, en los estudios por muestreo estratificado, es que las conclusiones pueden darse no sólo a nivel poblacional, sino que independientemente a nivel de cada estrato.

3.1.3. Muestreo Sistemático

Se forma una muestra sistemática seleccionando al azar una unidad y luego se seleccionan adicionalmente las unidades a intervalos igualmente espaciados (cada k unidades de la población, $k > 1$) hasta formar la muestra total.

Supongamos que estamos interesados en una población cuyos individuos han sido numerados desde 1 a N , y elegimos una muestra de tamaño n como sigue: entre las k primeras unidades, donde $k = N/n$ elegimos, en forma aleatoria, el elemento i — *sim*o, entre los k primeros elementos o unidades muestrales. A continuación, sistemáticamente se eligen los elementos $i + k$ que están k lugares después del i — *sim*o, y así sucesivamente hasta agotar los elementos disponibles de la lista, lo que ocurrirá al llegar al elemento que ocupa el lugar $i + (n - 1)k$.

Si existen k muestras posibles de n elementos cada una. Por ejemplo, de una población determinada por $N = 1000$ elementos deseamos elegir sistemáticamente 25 elementos, habrá $k = 1000/25 = 40$ elecciones diferentes posibles. Elegimos aleatoriamente un número comprendido entre 1 y 40 y a partir de éste, elegimos los 24 restantes de 40 en 40. Supongamos que fue sorteado el 15, entonces la observación que ocupa el 15° lugar será el primer elemento de la muestra, el segundo será el 55° (15 + 40), el tercero será el 95° (15 + 80) y así sucesivamente

hasta que el último elemento seleccionado será la observación que ocupe el lugar 975. La muestra estará conformado por los elementos : $15^\circ, 55^\circ, 95^\circ, \dots, 975^\circ$

Es más fácil de ejecutar este tipo de muestreo. Es también más preciso que el m.a.s. pues equivale a estratificar la población en n estratos de cada uno de los cuales se extrae una observación. En general, si se desea una muestra de tamaño n basta sortear una de las k columnas. Este tipo de muestreo es particularmente útil en muestreos exploratorios, en los que no se conoce exactamente la distribución de la población.

Este tipo de muestreo no es recomendable cuando el proceso generador de los datos tiene un comportamiento cíclico o estacional, el cuál puede hacer que las unidades seleccionadas coincidan con todos los máximos o todos los mínimos.

3.1.4. Muestreo por Conglomerados

Una muestra por conglomerados se obtiene identificando un conjunto de conglomerados que componen la población, aleatoriamente se selecciona un subconjunto de estos conglomerados y se hace un censo en cada uno de los seleccionados.

A veces, para estudios exploratorios, el muestreo probabilístico resulta excesivamente costoso y se acude a métodos no probabilísticos, aún siendo conscientes de que no sirven para realizar generalizaciones, pues no se tiene certeza de que la muestra extraída sea representativa, ya que no todos los sujetos de la población tienen la misma probabilidad de ser elegidos. En general se selecciona a los sujetos siguiendo determinados criterios procurando que la muestra sea representativa.

3.2. Muestreos no probabilísticos

En el caso de no contar con los supuestos que puedan garantizar la selección de la muestra de una manera aleatoria (no contar con un listado de la población o el no poder identificar con anticipación los elementos que conforman la población) y requerir de información para realizar análisis exploratorios, podemos utilizar otro tipo de muestreos llamados no probabilísticos. A continuación se describen algunos de ellos:

3.2.1. Muestreo por Cuotas

También denominado en ocasiones **accidental**. Se asienta generalmente sobre la base de un buen conocimiento de los estratos de la población y/o de los individuos más **representativos** o **adecuados** para los fines de la investigación. Mantiene, por tanto, semejanzas con el muestreo aleatorio estratificado, pero no tiene el carácter de aleatoriedad de aquel. En este tipo de muestreo se fijan unas "cuotas" que consisten en un número de individuos que reúnen unas determinadas condiciones, por ejemplo: 20 individuos de 25 a 40 años, de sexo femenino y residentes en una

determinada región. Una vez determinada la cuota se eligen los primeros que se encuentren que cumplan esas características. Este método se utiliza mucho en las encuestas de opinión.

3.2.2. Muestreo Intencional

Este tipo de muestreo se caracteriza por un esfuerzo deliberado de obtener muestras representativas mediante la inclusión en la muestra de grupos supuestamente típicos. Es muy frecuente su utilización en sondeos preelectorales de zonas que en anteriores votaciones han marcado tendencias de voto.

3.2.3. Muestreo Casual o Incidental

Se trata de un proceso en el que el investigador selecciona directa e intencionadamente los individuos de la población. El caso más frecuente de este procedimiento es utilizar como muestra los individuos a los que se tiene fácil acceso (los profesores de universidad emplean con mucha frecuencia a sus propios alumnos). Un caso particular es el de los voluntarios.

3.2.4. Bola de Nieve

Se localiza a algunos individuos, los cuales conducen a otros, y estos a otros, y así hasta conseguir una muestra suficiente. Este tipo se emplea muy frecuentemente cuando se hacen estudios con poblaciones "marginales", delincuentes, sectas, determinados tipos de enfermos, etc.

4. MÉTODOS DE ESTIMACIÓN

Para construir los estimadores se emplean métodos de estimación como los que se describen a continuación:

4.1 Método de momentos

El método de momentos fue propuesto por Karl Pearson al rededor de 1895, pensado en sus inicios en contexto descriptivo, analizando las distribuciones de probabilidad y aproximándolas al sistema de curvas que llevan su nombre. Posteriormente este concepto fue modificado por R.A. Fisher en 1920. El método consiste en estimar un parámetro de una distribución igualando sus momentos teóricos o poblacionales, si existen, con los correspondientes momentos muestrales.

Para mostrar este método es necesario definir el concepto de momento.

Momento Poblacional k-esimo:

- caso variable discreta

$$\mu^k = E[X^k] = \sum_{Rx} x^k p(x)$$

- caso variable continua

$$\mu^k = E[X^k] = \int_{-\infty}^{\infty} x^k f(x) dx$$

Momentos muestrales:

En ambos casos (v.a. discreta o continua)

$$m^k = \frac{1}{n} \sum_{i=1}^n x_i^k$$

El método de momentos supone que los momentos tanto poblacionales como muestrales son conocidos, y por lo tanto también la función de probabilidad.

A continuación se relacionan algunos de estos momentos poblacionales:

Distribución	$E[X]$	$V[X]$	Distribución	$E[X]$	$V[X]$
Bernoulli	p	pq	Gamma	$\alpha\beta$	$\alpha\beta^2$
Geométrica	$\frac{1}{p}$	$\frac{q}{p^2}$	Exponencial	β	β^2
Binomial	np	npq	Uniforme	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Poisson	λ	λ	Normal	μ	σ^2

Nota: Se podrían dar en función de λ , haciendo $\beta = \frac{1}{\lambda}$

Ejemplo

Encuentre los estimadores de los parámetros de la distribución normal a través del método de momentos.

Previamente sabemos que los parámetros de una variable con distribución normal son $E[X] = \mu$ y $V[X] = \sigma^2$ y que $V[X] = E[X^2] - E[X]^2$. Dada esta información el estimador de momentos se construye de la siguiente manera:

$$\mu^1 = m^1$$

$$\mu^2 = m^2$$

Aplicando el método:

$$\mu^1 = E[X] = m^1$$

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i$$

Finalmente,

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$$

Para estimar σ^2 , se realiza el siguiente procedimiento, usando $\mu^1 = m^1$ y $\mu^2 = m^2$.

$$V[X] = E[X^2] - E[X]^2 = \mu^2 - (\mu^1)^2$$

entonces igualamos estos dos momentos poblacionales con sus respectivos momentos muestrales quedando la igualdad

$$\begin{aligned} V[X] &= \mu^2 - (\mu^1)^2 \\ &= m^2 - (m^1)^2 \\ &= \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 \end{aligned}$$

podemos representar la varianza por σ^2 y obtenemos

$$\widehat{\sigma^2} = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2$$

y obtenemos el estimador de la varianza:

$$\begin{aligned} \widehat{\sigma^2} &= \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 \\ \widehat{\sigma^2} &= \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 - \bar{x}^2 + \bar{x}^2 \\ &= \frac{1}{n} \sum_{i=1}^n x_i^2 - 2\bar{x}^2 + \bar{x}^2 \\ &= \frac{1}{n} \sum_{i=1}^n x_i^2 - \frac{2\bar{x} \sum_{i=1}^n x_i}{n} + \frac{n\bar{x}^2}{n} \\ &= \frac{1}{n} \left(\sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + n\bar{x}^2 \right) \\ \widehat{\sigma^2} &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \end{aligned}$$

En resumen los estimadores de momentos para los parámetros de la distribución normal son:

$$\begin{aligned} \hat{\mu} &= \frac{1}{n} \sum_{i=1}^n x_i = \bar{x} \\ \widehat{\sigma^2} &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \end{aligned}$$

Ejemplo

A partir de ellos y mediante la obtención de una muestra aleatoria por ejemplo :630, 650, 710, 750, 790, 820, 860 y 910 se pueden estimar los parámetros por método de momentos con los siguientes resultados:

$$\hat{\mu} = 765$$

$$\widehat{\sigma^2} = 8550$$

4.2 Método de máxima verosimilitud

Uno de los mejores métodos para obtener un estimador puntual de un parámetro es el método de máxima verosimilitud o de máxima probabilidad. Esta técnica fue desarrollada en 1920 por el estadístico británico Sir R.A. Fisher. El estimador será el valor del parámetro que maximice la función de verosimilitud $L(\theta)$.

La función de verosimilitud $L(\theta)$ corresponde a la función de distribución conjunta de variables aleatorias independientes con igual función de distribución. Estas variables aleatorias corresponden a las variables que conforman la muestra.

$$L(\theta) = f(x_1, \theta) \cdot f(x_2, \theta) \cdot f(x_3, \theta) \dots f(x_n, \theta)$$

El objetivo del método será encontrar el valor del parámetro que maximice la probabilidad conjunta.

El método supone el conocimiento de la función de distribución de probabilidad de la variable en estudio. Por ejemplo:

En la unidad anterior tratamos en caso de la distribución normal cuya función de distribución de probabilidad está dada por :

$$f(x_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right)$$

La función de verosimilitud estará dada por:

$$L(x_1, x_2, \dots, x_n; \mu, \sigma^2) = f(x_1; \mu, \sigma^2) \dots f(x_n; \mu, \sigma^2)$$

Esta función se puede escribir como :

$$L(x_1, \dots, x_n; \mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right)$$

$$L = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left(\sum_{i=1}^n \frac{-1}{2\sigma^2}(x_i - \mu)^2\right)$$

$$L = (2\pi\sigma^2)^{-n/2} \exp\left(\frac{-1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right)$$

El método consiste en encontrar el valor del parámetro que maximice esta función para lo cual procedemos a derivar L parcialmente con respecto a μ .

Este proceso presenta algunas dificultades de cálculo que son atenuadas mediante la premisa de que el máximo de la función L corresponde a los mismos máximos de la función $\ln(L)$, la cual es más sencilla de derivar. Este procedimiento es posible debido a que la función L es creciente. Convertimos L en $\ln(L)$

$$\ln(L) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

Al derivar parcialmente $\ln(L)$ con respecto a μ tenemos:

$$\frac{\partial \ln(L)}{\partial \mu} = \frac{2}{2\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0$$

De esta igualdad se despeja el parámetro de interés

$$\frac{2}{2\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0$$

$$\sum_{i=1}^n x_i - n\mu = 0$$

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$$

En el caso de la estimación de σ^2 , se deriva $\ln(L)$ parcialmente con respecto a σ^2 , se iguala a cero el resultado obtenido y por último se despeja σ^2 . Verifique que el estimador de máxima verosimilitud para la varianza es igual a:

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

Algunas propiedades de la función $\ln(x)$

$$\begin{aligned} \ln(xy) &= \ln(x) + \ln(y) & \ln(x/y) &= \ln(x) - \ln(y) \\ \ln(x^n) &= n \ln(x) & \ln(e^x) &= x \\ \partial \ln(x) &= \frac{1}{x} \end{aligned}$$

5. Propiedades deseables en los estimadores

Anteriormente se describieron los dos métodos para la construcción de estimadores, Estos métodos generan diversas alternativas dentro de los cuales debemos seleccionar los mejores. Para realizar dicha clasificación debemos examinar sus principales propiedades como son la INSESGADEZ, CONSISTENCIA, EFICIENCIA y SUFICIENCIA, entre otras.

Un indicador que mide la calidad de un estimador $\hat{\theta}$ se denomina Error Cuadrático Medio (ECM) y se define de la siguiente manera:

Error Cuadrático Medio

Se define como **Error Cuadrático Medio** (ECM) el valor esperado de la diferencia entre el estimador y el parámetro, al cuadrado.

$$\begin{aligned} ECM[\hat{\theta}] &= E[\hat{\theta} - \theta]^2 \\ &= E[\hat{\theta}^2 - 2\hat{\theta}\theta + \theta^2] \\ &= E[\hat{\theta}^2] - 2\theta E[\hat{\theta}] + \theta^2 \\ &= E[\hat{\theta}^2] - E[\hat{\theta}]^2 + E[\hat{\theta}]^2 - 2\theta E[\hat{\theta}] + \theta^2 \\ &= V[\hat{\theta}] + \left[\theta - E[\hat{\theta}] \right]^2 \\ &= V[\hat{\theta}] + \text{Sesgo}^2 \end{aligned}$$

Insesgadez

Se dice que el estimador $\hat{\theta}$ que es función de los datos contenidos en una muestra, es un estimador insesgado del parámetro θ , si $E[\hat{\theta}] = \theta$, para todos los posibles valores de θ (Canavos(1988))

Suponga que una muestra aleatoria simple de X_1, X_2, \dots, X_n procede de una población con $E[X] = \mu$, el parámetro de interés. Probar que la media muestral \bar{X} es siempre un estimador insesgado del parámetro media poblacional μ .

$$\begin{aligned}
E[\bar{X}] &= E\left[\frac{1}{n}(X_1 + X_2 + \dots + X_n)\right] \\
&= \frac{1}{n}E[X_1 + X_2 + \dots + X_n] \\
&= \frac{1}{n}[E[X_1] + E[X_2] + \dots + E[X_n]] \\
&= \frac{1}{n}[\mu + \mu + \dots + \mu] \\
&= \frac{1}{n}n\mu = \mu
\end{aligned}$$

Por tal razón $\hat{\mu} = \bar{x}$ es un estimador insesgado de μ

Ejemplo

Para una muestra aleatoria simple de X_1, X_2, \dots, X_n procede de una población con $V[X] = \sigma^2$, el parámetro de interés. Determinar si el siguiente estimador es insesgado o no:

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

Entonces:

es necesario tener en cuenta que:

$$V[X] = \sigma^2 = E[X^2] - E[X]^2$$

$$V[\bar{X}] = \frac{\sigma^2}{n} = E[\bar{X}^2] - E[\bar{X}]^2$$

a partir de estos resultados podemos tener

$$\begin{aligned}
E[\hat{\sigma}_n^2] &= E\left[\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2\right] \\
&= \frac{1}{n}E\left[\sum_{i=1}^n (x_i - \bar{x})^2\right] \\
&= \frac{1}{n}E\left[\sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + n\bar{x}^2\right] \\
&= \frac{1}{n}E\left[\sum_{i=1}^n x_i^2 - 2n\bar{x} \frac{\sum_{i=1}^n x_i}{n} + n\bar{x}^2\right] \\
&= \frac{1}{n}E\left[\sum_{i=1}^n x_i^2 - 2n\bar{x}^2 + n\bar{x}^2\right] \\
&= \frac{1}{n}E\left[\sum_{i=1}^n x_i^2 - n\bar{x}^2\right] \\
&= \frac{1}{n}\left[\sum_{i=1}^n E[x_i^2] - nE[\bar{x}^2]\right]
\end{aligned}$$

reemplazando por los respectivos valores tenemos

$$\begin{aligned}
 E[\widehat{\sigma}_n^2] &= \frac{1}{n} \left[\sum_{i=1}^n [\sigma^2 + \mu^2] - n \left[\frac{\sigma^2}{n} + \mu^2 \right] \right] \\
 &= \frac{1}{n} \left[[n\sigma^2 + n\mu^2] - \left[n\frac{\sigma^2}{n} + \mu^2 \right] \right] \\
 &= \frac{1}{n} [n\sigma^2 - \sigma^2] \\
 &= \frac{1}{n} [(n-1)\sigma^2] \\
 &= \frac{n-1}{n} [\sigma^2]
 \end{aligned}$$

por lo tanto $\widehat{\sigma}_n^2$ es un estimador sesgado de σ^2

Consistencia

Se dice que $\widehat{\theta}$ es un **estimadores consistente** de θ si al aumentar el tamaño de la muestra se convierte en un estimador insesgado de θ

(Sarabia 2007)

Ademas,

Sea $\widehat{\theta}$ el estimador de un parámetro θ y sea $\widehat{\theta}_1, \widehat{\theta}_2, \dots, \widehat{\theta}_n$ una secuencia de estimadores que representan a $\widehat{\theta}$ con base en muestras de tamaño 1, 2, ..., n, respectivamente. Se dice que $\widehat{\theta}$ es un estimador consistente para θ si:

$$\lim_{n \rightarrow \infty} P(|\widehat{\theta} - \theta| \leq \epsilon) = 1$$

para todos los valores de θ y $\epsilon > 0$

Ejemplo

Como ejemplo de estimador consistente podemos tomar el estimador sesgado

$$\widehat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

Como se demostró:

$$E[\widehat{\sigma}_n^2] = \frac{n-1}{n} \sigma^2$$

Sin embargo cuando el tamaño de muestra n tiende a infinito el factor : $(n-1)/n$ tiende a 1 y de esta manera el estimador $\widehat{\sigma}_n^2$ se convierte en un estimador insesgado

Mínima varianza o Eficiencia

Sean $\hat{\theta}_1$ y $\hat{\theta}_2$ dos estimadores insesgados del parámetro θ :

- Se dice que el estimador $\hat{\theta}_1$ es más eficiente que el estimador $\hat{\theta}_2$ si

$$V[\hat{\theta}_1] < V[\hat{\theta}_2]$$

- Se define la eficiencia relativa de un estimador con respecto a otro como el cociente de las varianzas:

$$\frac{V[\hat{\theta}_2]}{V[\hat{\theta}_1]}$$

Existe la **cota de Cramer Rao**, que permite estimar un valor mínimo de la varianza. En caso de que algún estimador tenga como varianza este valor decimos que este estimador es eficiente.

$$V[\hat{\theta}] \geq \frac{1}{nE\left[\left(\frac{\partial \ln(f(X, \theta))}{\partial \theta}\right)^2\right]}$$

Ejemplo

Para determinar si el estimador \bar{x} de μ en una población con distribución normal es eficiente, es necesario calcular la cota de Cramer Rao.

$$\begin{aligned} f(x, \mu, \sigma^2) &= \left(\frac{1}{2\pi\sigma^2}\right)^{1/2} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \\ \ln(f(x, \mu, \sigma^2)) &= -\frac{1}{2} \ln(2\pi\sigma^2) - \frac{(x-\mu)^2}{2\sigma^2} \\ \frac{\partial \ln(f(x))}{\partial \mu} &= \frac{2(x-\mu)}{2\sigma^2} \\ \left[\frac{\partial \ln(f(x))}{\partial \mu}\right]^2 &= \frac{(x-\mu)^2}{\sigma^4} \\ nE\left[\frac{\partial \ln(f(x))}{\partial \mu}\right]^2 &= n\left[\frac{E(x-\mu)^2}{\sigma^4}\right] \\ nE\left[\frac{\partial \ln(f(x))}{\partial \mu}\right]^2 &= n\frac{\sigma^2}{\sigma^4} = \frac{n}{\sigma^2} \\ \frac{1}{nE\left[\frac{\partial \ln(f(x))}{\partial \mu}\right]^2} &= \frac{\sigma^2}{n} \end{aligned}$$

Como resultado obtenemos que la cota de Cramer-Rao para un estimador de μ es igual a σ^2/n , valor que corresponde a la varianza de \bar{x} . De este resultado podemos concluir que \bar{x} es el estimador de mínima varianza para μ en el caso de la distribución normal.

Suficiencia

Se dice que un estimador $\hat{\theta}$ es un estimador suficiente del parámetro θ , si la distribución condicional de la muestra ($L(x, \theta)$) no depende de θ (Sarabia(2007))

Otra forma de establecer si un estimador es suficiente es:

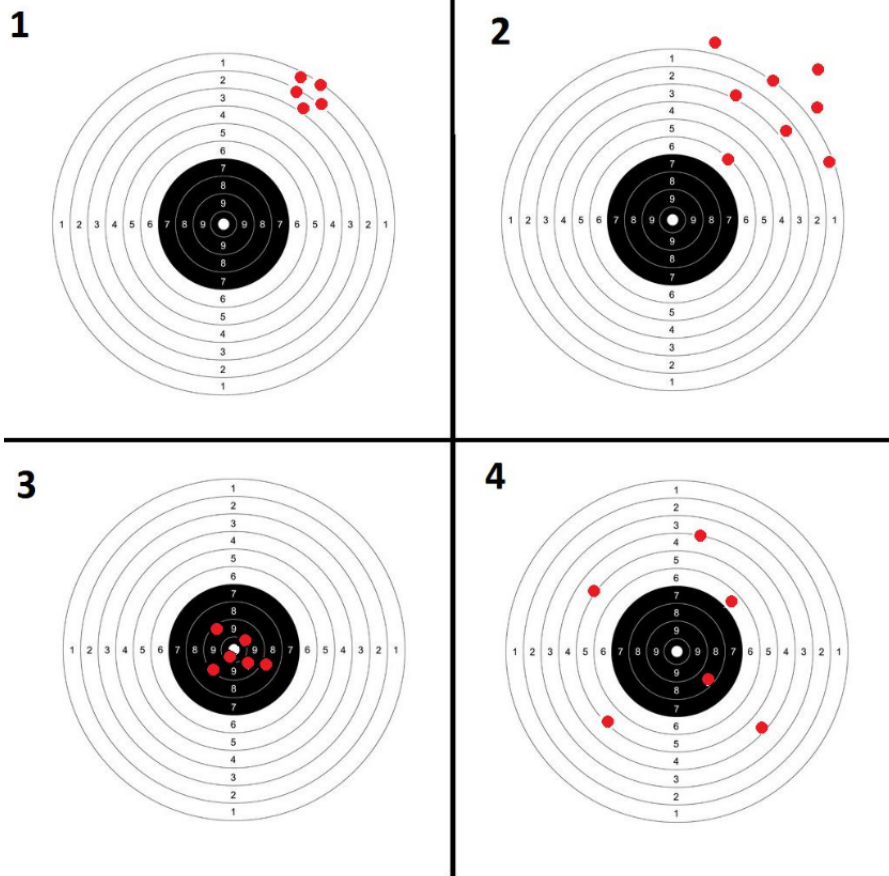
Criterio de factorización de Fisher

Sea $L(x, \theta)$ la función de probabilidad de la muestra. El estimador $\hat{\theta}$ es suficiente para la estimación de θ si solo si existen funciones g y h tales que:

$$L(x, \theta) = g(x)h(\hat{\theta}(x))$$

Con g una función no negativa que solo depende de la información de la muestra (x) y h que es una función no negativa que depende de $\hat{\theta}$ y θ . (Sarabia(2007)).

Criterios para seleccionar un buen estimador



Cuando se escoge un estimador se debe elegir aquel que genere las mejores condiciones de estimación. Pensemos que en el centro de la diana está el parámetro objetivo. Si elegimos un estimador que proporcione resultados similares al cuadrante 1 estaremos optando por un estimador

que tiene una varianza pequeña, pero está alejado del centro estimador sesgado. En el caso del cuadrante 2, el estimador tienen una varianza grande, pero además es sesgado. En el cuadrante 3 tenemos un estimador insesgado - en promedio damos en el centro - pero presenta una varianza grande. Finalmente el estimador representado por el cuadrante 4, presenta las mejores condiciones, tiene varianza pequeña y es insesgado.

También es posible elegir un estimador sesgado, al que sea posible estimar su sesgo y ser corregido o un estimador insesgado pero consistente, es decir que en principio sea sesgado pero que si aumentamos el tamaño de la muestra se torne insesgado.

Ejercicios propuestos

- Si X_1, X_2, \dots, X_n constituye una muestra aleatoria, obtenga la función de verosimilitud para las siguientes distribuciones:
 - Poisson con parametro λ
 - Uniforme con parametros (a, b)
 - Gamma con parametros α, β
 - Weibull con parametros α, β
- Estime los parámetros enunciados en el punto [1] mediante el método de máxima verosimilitud
- Estime los parámetros enunciados en el punto [1] mediante el método de momentos
- Demuestre que el estimador

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

es un estimador insesgado

- A partir de una muestra de tamaño 3 procedente de una población de Bernoulli de parámetro p se consideran los siguientes estimadores:

$$\hat{p}_1 = \frac{X_1 + X_2 + X_3}{3}$$

$$\hat{p}_2 = \frac{X_1 + X_2 + 1}{3}$$

$$\hat{p}_3 = \frac{X_1 + X_2 + 2X_3}{4}$$

- Hallar el sesgo, la varianza y el error cuadrático medio de los tres estimadores.
- Cuál de los tres estimadores es preferible?, justifique su respuesta.
- Si para la selección de estimadores exigimos que sea insesgado, ¿cuál es ahora preferible?.

6. Considere la función de densidad de probabilidad:

$$f(x) = c(1 + \theta x) \text{ para } -1 \leq x \leq 1$$

- (a). Determine el valor de c
- (b). Encuentre el estimador de momentos para θ
- (c). Muestre que si $\hat{\theta} = 3\bar{X}$, es un estimador insesgado de θ
- (d). Encuentre el estimador de máxima verosimilitud para θ