**Estrategia didáctica 2.3.1.1. Población y muestra**

**Comentario**: *Con esta estrategiaa, consiste exclusivamente en una serie de preguntas, se inicia el estudio conceptual de la estadística. Se revisan brevemente los conceptos de muestreo, población y otros se dejan a la crítica y discusión del alumno*.

**I.** Cuando se realiza un estudio, en muchas ocasiones es necesario establecer una hipótesis que intente explicar el fenómeno de estudio. La hipótesis se puede considerar como un enunciado que puede ser falso o verdadero y que es una primera explicación del fenómeno de interés. Por ejemplo, se puede decir que la televisión está relacionada con el alto porcentaje de alumnos reprobados. Si un alumno invierte mucho tiempo en ver televisión, esto incidirá en su desempeño escolar y sus calificaciones serán bajas. Esta creencia es una hipótesis que se debe contrastar para decidir si esto es como se plantea. Debe ser apropiado determinar a qué tipo de personas se les entrevistará para decidir si la hipótesis está siendo contrastado o no es así. Por ejemplo, se pueden entrevistar o encuestar a varias personas de distintas edades y preguntarles cuánto tiempo ven televisión y su desempeño en la escuela que pudiera ser medido con la calificación promedio que tienen en todos sus cursos. Pero es claro que no todas las personas ven el mismo tiempo televisión, pues es muy posible que para distintas edades, el tiempo que ven televisión es distinto. Por esta razón, debemos especificar el tipo de personas que deberemos encuestar para que la información que obtengamos sea relevante al problema. Esta se le llama la población objetivo.

Podemos afirmar que toda hipótesis que no se contraste, pasa a ser un prejuicio. Pero no siempre tenemos las posibilidades de contrastar toda hipótesis porque es muy probable que no tengamos los medios para hacerlo. En muchas ocasiones es necesario revisar toda la población de interés, lo que en la mayoría de las ocasiones es imposible y caro y por tanto el problema es difícil de resolver. Sin embargo, podemos darnos una idea del estado que guarda el problema de interés si extraemos una muestra, es decir, un subconjunto de la población, a la cual le realizaremos algunos estudios y los resultados que obtengamos los inferiremos a la población de estudio. Sin embargo, debemos tener algunas precauciones porque si la muestra fue mal seleccionada, es posible que nuestros resultados no se apliquen a la población de interés porque la muestra no guarda relación alguna con aquella. Por ejemplo, si entrevistamos a las amas de casa para preguntarles cuál es su programa de TV favorito, cometeríamos un error si a partir de estos resultados concluyamos los gustos televisivos de toda la población, o de todos los niños menores de 10 años, porque los gustos cambian con la edad. A este tipo de errores, en los que se ha considerado una muestra que no tienen las características de la población objetivo, se le llama sesgo de selección. Existen varios tipos de sesgo de selección, que se verán más adelante:

Una muestra representativa de una población se toma para:

* inferir características de la población (inferir es una afirmación relativa a poblaciones estadísticas, efectuada a partir de ciertas observaciones con determinada medida de la incertidumbre – Cox (1958))
* tomar decisiones

 Definimos la población como un conjunto de elementos, objetos o seres vivos con características comunes y muchas no comunes. (Por ejemplo los *estudiantes: adolescentes, mexicanos, 1 a 4 años de estudio; alimentación, tipo de vivienda, medio de transporte, ambiente familiar)*

Se tienen en particular las poblaciones de interés, que es la población objetivo y la población muestreada: en general ambas no siempre coinciden porque la primera consta de todos los objetos o personas que quisiéramos que fuera posible encuestar, mientras que la segunda es la población que está en posibilidades de ser encuestada, por ejemplo, si deseamos hacer un estudio de los alumnos que ven televisión, la población objetivo consiste de todos aquellos alumnos que están inscritos en el CCH, pero la población de la que será posible tomar una muestra es de aquella de la que los alumnos asisten regularmente a la escuela, pues existen varios que a pesar de estar inscritos, no asisten a clase, o han dejado de asistir por diversas razones o que simplemente no desean contestar la encuesta que aplicaremos. A veces, se dispone de una lista de los elementos que vamos a muestrear de una población, a esta lista se le llama marco de muestreo. No siempre tenemos un marco de muestreo, sobre todo cuando vamos a tomar muestras de peces o de frutas de un embarque, pero a veces si disponemos de uno de ellos, como cuando tenemos una lista de los alumnos inscritos en una escuela o del padrón de personas que pueden votar en una elección.

Por lo anterior, no siempre podemos tener a la población de interés disponible en el momento de hacer una encuesta. Esto propicia diversos sesgos o errores de selección, los cuales ocurren cuando una parte de la población objetivo no esta en la población muestreada.

Formas en que se presenta el sesgo de selección:

1. Cuando los procedimientos de selección de la muestra dependen de la característica asociada a las propiedades de interés *(mujeres golpeadas)*
2. Muestra de juicio. Cuando se seleccionan a las personas que creemos que pueden conformar una muestra representativa*. (Por ejemplo seleccionar a los alumnos que creemos que ven televisión y recorremos el Colegio de manera azarosa seleccionando a quien nos parezca.)*
3. Errores en la especificación de la población objetivo. *(No ser claros en el tipo de población a la que vamos a seleccionar, por ejemplo si deseamos encuestar a migrantes, no definir claramente lo que significa que alguien sea un migrante. )*
4. No incluir a toda la población objetivo en el marco de muestreo. *(Por ejemplo, todos los días se crean empresas nuevas y quiebran otras. La información no se actualiza de manera inmediata.)*
5. Sustitución de un elemento por otro. Cuando se selecciona un elemento que no se encontró y se decide encuestar a uno cercano.
6. Ausencia de respuesta.
7. Muestra integrada por voluntarios *(Las personas que van a los hospitales por alguna enfermedad y que son a los que encuestamos)*

Existen otros errores de sesgo. Entre ellos están los errores de medición. Estos ocurren cuando el método de medición (encuesta o experimento) no permite la obtencion completa de la informacion.

### Formas en que se presenta el sesgo de medición:

1. las personas no dicen la verdad *( No siempre responderán si pagan o no pagan impuestos.)*
2. Las preguntas no se comprenden. Cuando están tan elaboradas que la redacción confunde al encuestado.
3. La gente olvida sucesos. *(Es difícil recordar cuando se vio un programa de televisión específico.)*
4. Distintas respuestas a diferentes encuestadores. La gente responderá de manera distinta a un encuestador dependiendo de cómo viste o si se trata de una mujer o de un hombre.
5. La respuesta se orienta a lo que se cree que el entrevistador quiere oir. (Porque se quiere impresionarlo.)
6. El encuestador puede sesgar la respuesta. Pues es posible que el encuestador sugiera respuesta según el tono de voz o que esté convencido del valor de una respuesta y trate de inducirla.
7. Las personas dan sentidos distintos a las palabras. Por ejemplo, la palabra “hostigar” puede ser interpretada de manera distinta por una mujer.
8. El orden y la formulación de las preguntas pueden alterar las respuestas.

Así pues, existen dos tipos de errores cuando se realiza una encuesta: errores de muestreo que son los que se obtienen por tomar una muestra, y los errores de no muestreo que los que se obtienen por sesgo de selección. La toma de una muestra se realiza cuando se quiere obtener información rápida y precisa para tomar decisiones acerca del estado de la población de interés, pero esto implica que las características de la muestra sean lo bastante parecidas con las características de la población para que las decisiones sean lo más adecuadas posible. La razón de ello, es que los datos obtenidos de la encuesta son los que harán que se pueda contrastar la hipótesis, es decir, que la apoyen o que la rechacen definitivamente.

Esto se resume en el siguiente diagrama:

 inferencia

 muestra población

Existe la posibilidad de equivocarse cuando se hace una inferencia, pero la estadística procura que este error sea el más pequeño posible. Esto puede entenderse diciendo que las gráficas o características que observamos en la muestra son bastante cercanas a las características de la población o a las gráficas que describen a la población. Si esto sucede, es muy probable que el problema de interés se resuelva con muchas posibilidades de que la solución sea correcta, pero no debemos olvidar que a pesar de ello, podemos tener conclusiones equivocadas que incidirán en el análisis de la hipótesis.

Con relación a las hipótesis, estas pueden tener varias posibilidades que se pueden dar cuando se están contrastando al tomar una muestra, por ejemplo:

1. La hipótesis puede ser verdadera, y los datos la apoyan.
2. La hipótesis puede ser verdadera, pero los datos la rechazan.
3. La hipótesis puede ser falsa, y los datos la apoyan.
4. La hipótesis puede ser falsa, pero los datos la rechazan.

Por ejemplo, cuando se conoce a una persona y se quiere establecer amistad con ella, se empieza siempre haciendo hipótesis acerca de sus gustos o de su carácter y si estos son compatibles con los de uno mismo, es posible que se pueda lograr algo más. La hipótesis sería que “la muchacha es agradable”, que puede ser una entre muchas hipótesis, pues no sólo se plantea una, pero siempre es importante trabajar con el menor número posible de hipótesis pues de lo contrario la encuesta sería muy larga.

 En el ejemplo que se comenta, para (1) la muchacha es agradable y las salidas que se han tenido con ella apoyan la hipótesis. Esto significa que decidiremos seguirla frecuentando; para (2) también puede ser que la muchacha sea agradable pero las pláticas que se han tenido con ella indican lo contrario o las veces que se ha salido con ella dicen lo contrario, decidiremos pues, dejarla de ver, pero esto es un error porque en verdad era agradable, a este error se le llama error de tipo I que consiste en falsar cosas verdaderas; en el caso (3), se tiene que la muchacha no es agradable pero después de platicar con ella creemos que lo es, decidiremos seguir saliendo con ella cuando en realidad no es como creemos, a este caso se le llama error de tipo II, que consiste en creer cosas falsas, y finalmente el caso (4) se tiene que la muchacha no es agradable y las pláticas y salidas que hemos tenido con ella nos indican que así es, por lo que decidiremos no seguir frecuentándola.

Debe notarse que en todos los casos, realizar un estudio con hipótesis y diseño de encuesta, implican la toma de una decisión: en el caso del ejemplo anterior, se decidirá si se continúa viendo a la muchacha o no se continúa con ella. En todo caso, no debe olvidarse que las decisiones pueden ser equivocadas cuando se ha tomado una muestra sesgada ya sea por selección o medición. En nuestro ejemplo, el muestreo se hace cuando se platica con la muchacha o cuando se sale con ella.

La realización de un estudio, debe tener fundamentos metodológicos, es decir, cómo hacer el diseño o la encuesta y cómo aplicarla par contrastar la hipótesis de la mejor manera posible. Esta metodología que debemos emplear en la elaboración de estudios es la siguiente:

“Para efectuar el trabajo de la estadística aplicada, se requieren también conocimientos y toma de posición respecto a aspectos epistemológicos. Es común que muchos estadísticos consideren a su profesión como una de “expertos en inferencia” o como “metodología de investigación”. Esto por que se tiene una forma de estudiar los fenómenos con mucha variabilidad, independientemente que se refieran a personas, plantas, animales, piezas de metal, familias o empresas, etc.

 Se considera que el uso de la estadística procura mas objetividad en la adquisición de conocimiento o la toma de decisiones relativas a ese tipo de fenómenos.

 Los datos, la información numérica, no se puede colectar sin ideas preexistentes sobre lo que es relevante y lo que no; sobre las formas de observar, medir o experimentar (diseño); sobre las formas de modelar o de analizar. Así, las conclusiones están influenciadas por esas ideas preexistentes. Además, y de forma muy importante, por los métodos estadísticos conocidos y disponibles a los investigadores. Así la estadística aplicada se convierte en un “arte, donde se conjugan los aspectos teóricos de la estadística y las disciplinas fácticas, con la experiencia y habilidades de los investigadores.” Dentro de la metodología, para la estadística resalta el concepto de validez externa.

Todas las investigaciones se llevan a cabo, con un grupo de elementos (la o las muestras), con el objeto de generalizar los resultados de ellas a un grupo mucho mayor de elementos (la o las poblaciones). Es importante resaltar que es necesario definir adecuadamente estos conjuntos (tanto muestras como poblaciones). Para hacer las extrapolaciones de muestra a población, se requiere tener una muestra representativa de la población. Sin embargo, existen situaciones de investigación donde la población es infinita, sin ubicación precisa en tiempo y espacio; por ejemplo todos los seres humanos adultos con leucemia, en un estudio sobre diversos tratamientos, o toda lo progenie posible de una planta con reproducción vegetativa.. La dificultad estriba en que usualmente, se dispone de un grupo de individuos, no seleccionados aleatoriamente de la población, son los disponibles; por ejemplo los que llegan a un hospital en un periodo; o las plantas localizadas en un sitio en una fecha dada. En este caso es necesario precisar el tipo de población para la cual la muestra es representativa. Ahora se parte de la muestra para definir o especificar la población. En ambos casos (de población a muestra o viceversa), se debe siempre valorar si la muestra reproduce o refleja los aspectos esenciales (según el objetivo del estudio) de la población. Si se considera que si se reproducen esos aspectos, se dice que hay validez externa.En importante hacer énfasis que las conclusiones de inferencias estadísticas se referirán a las poblaciones.” (I. Méndez.)

**II.** Ejercicios

1. Define lo que entiendas por “muestra” y por “población”.

Población es un conjunto total de elementos y muestra es una selección de cierto número de esos elementos.

1. Define, con tus propias palabras lo que entiendas por “aleatorio”.
2. ¿Qué entiendes por una variable estadística?
3. Da 2 ejemplos de poblaciones en las que a) Todos los elementos pueden ser medidos; b) Sólo una parte de los elementos puede ser medida. Piensa en la variable del problema.
* Personas con obesidad en Coahuila.
* Personas de saltillo que estudian

b)

* Personas con obesidad en Coahuila que realizan deporte.
* Personas de saltillo que estudian y trabajan
1. ¿Qué entiendes acerca del concepto “muestra representativa”? Dilo con tus propias palabras y da un ejemplo de ello. Da un ejemplo de una muestra no representativa.

Es un cierto número de una población que representa a toda la población.

1. ¿Cómo seleccionarías una muestra aleatoria de 70 mujeres mexicanas mayores de 12 años de todo el país?

Numerando todos los datos y utilizando la formula n. de datos Ran#= y según el número que aparezca elegir el dato que tenga ese número y hacer eso 70 veces

1. Supongamos que tienes la siguiente hipótesis: “La acupuntura es efectiva para curar dolor de espalda”. Explica los errores de tipo I y de tipo II, para esta hipótesis, breve pero precisamente.

Tipo I: La mayoría de las personas creen que la acupuntura no cura el dolor de espalda porque los doctores lo niegan

Tipo II: Las personas creen en que la acupuntura cura el dolor de espalda ya que lo comprobaron.

1. Supongamos que tienes la siguiente hipótesis: “Los dulces curan la hepatitis”. Explica los errores de tipo I y de tipo II, para esta hipótesis, breve pero precisamente.

Tipo I: Los doctores dicen que los dulces no curan la hepatitis, por lo que las personas con hepatitis no los consumen para curarse..

Tipo II: Las personas comen dulces porque piensan que curan la hepatitis, pero no lo pudieron comprobar.

1. Diseña una encuesta en la que decidas si una muchacha te conviene o no te conviene. ¿Qué le preguntarías? ¿en qué orden lo harías?
2. Da un problema donde una muestra tenga un sesgo de selección y un sesgo de medición.
* Personas que utilizan el transporte público.

Preguntar en una parada de transporte público.

1. Explica cómo se puede en una encuesta relativa al alcoholismo en los adolescentes cometer cada uno de los sesgos de medición.

Explica cómo se puede en una encuesta relativa a las preferencias sexuales de los jóvenes cometer cada uno de los sesgos de selección.

* Cuando los procedimientos de selección de la muestra dependen de la característica asociada a las propiedades de interés: Todos los hombres se interesan por las mujeres y todas las mujeres por los hombres.
* Muestra de juicio. Cuando se seleccionan a las personas que creemos que pueden conformar una muestra representativa. Preguntar solamente a nuestros amigos cercanos y decir que todos son heterosexuales.
* Errores en la especificación de la población objetivo. No definir que son los homosexuales o los heterosexuales.
* No incluir a toda la población objetivo en el marco de muestreo. Cada día se descubre una persona heterosexual.
* Sustitución de un elemento por otro. Cuando se selecciona un elemento que no se encontró y se decide encuestar a uno cercano. Entrevistar a mi hermano porque alguna persona no me quiso responder.
* Ausencia de respuesta. No querer responder.
* Muestra integrada por voluntarios Solamente las personas heterosexuales fueron entrevistadas.
1. Explica qué es un censo y busca cuáles son las diferencias con una muestra.

Un censo es cuando se agregan TODOS los datos obtenidos y una muestra solo son tomados algunos de esos datos.

**Guardar las preguntas** con el nombre **nombre-apellido.E2.3.1.1Población-muestra-grupo.doc**