

Muestreo y estadística descriptiva

Universidad de Guayaquil

October 23, 2018

**“Existen tres clases de mentiras: mentiras, viles mentiras y estadísticas”.
Se ha dicho que “las cifras no mienten, pero los mentirosos también usan cifras”**

- Los datos muestrales deben reunirse de una forma adecuada, como a través de un proceso de selección aleatoria.
- Si los datos muestrales no se reúnen de forma apropiada, resultarán tan inútiles que ninguna cantidad de tortura estadística podrá salvarlos.

Muestras de respuesta voluntaria (o muestra autoseleccionada)

Es aquella en que los propios sujetos deciden ser incluidos.

- Encuestas por Internet, donde los sujetos deciden si responden o no.
- Encuestas por correo, donde los sujetos deciden si responden o no.
- Encuestas telefónicas, donde anuncios televisivos, de radio o de periódicos le piden que llame voluntariamente a un número especial para registrar su opinión.

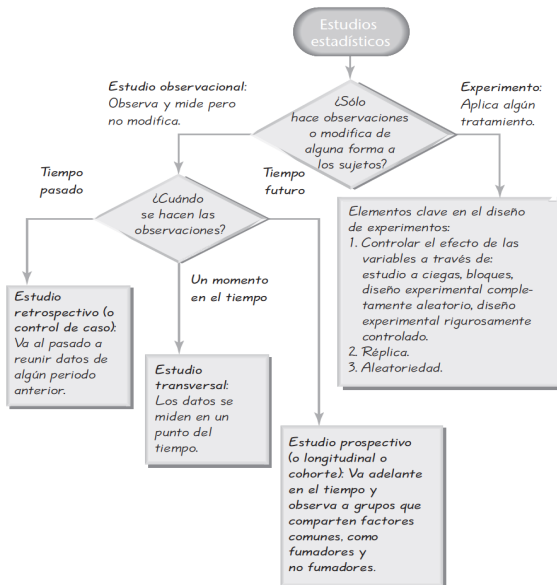
NOTA

Con este tipo de muestras de respuesta voluntaria sólo se logran conclusiones válidas sobre el grupo de gente específico que decidió participar; aunque una práctica común consiste en afirmar o sacar conclusiones incorrectas sobre una población más grande. Desde un punto de vista estadístico, una muestra de este tipo es defectuosa y no debe usarse para hacer afirmaciones generales sobre una población más grande.

Los métodos estadísticos se rigen por los datos. Por lo regular obtenemos datos de dos fuentes distintas: los estudios observacionales y los experimentos.

- En un estudio observacional, vemos y medimos características específicas, pero no intentamos modificar a los sujetos que estamos estudiando.
- En un experimento aplicamos algunos tratamientos y luego procedemos a observar sus efectos sobre los sujetos (en los experimentos, a los sujetos se les denomina unidades experimentales).

Fuente de los Datos (II)



Diseño de experimentos (I)

1. Control de los efectos de las variables.

Estudio a ciegas

Una técnica en que el sujeto no sabe si está recibiendo un tratamiento o un placebo

Diseño de bloques aleatorio

Utilice si está realizando un experimento para probar uno o más tratamientos diferentes. Forme bloques de sujetos con características similares. Asigne los tratamientos de manera aleatoria a los sujetos dentro de cada bloque.

Diseño experimental completamente aleatorio

Es completamente aleatorio, los sujetos se asignan a distintos grupos de tratamiento mediante un proceso de selección aleatoria.

Diseño rigurosamente controlado

Los sujetos se eligen cuidadosamente, de manera que quienes reciban cada tratamiento sean similares en los aspectos que son importantes para el experimento.

2. Réplica y tamaño de la muestra. Las muestras deben ser lo suficientemente grandes para que el comportamiento errático, que es característico de muestras muy pequeñas, no disfrace los efectos verdaderos de los distintos tratamientos.

La repetición de un experimento sobre un grupo de sujetos suficientemente grande se conoce como réplica

Nota

Utilice una muestra de un tamaño que sea lo bastante grande para distinguir la verdadera naturaleza de cualquiera de los efectos, y obtenga la muestra utilizando un método apropiado, como uno basado en la aleatoriedad.

3. Aleatoriedad y otras estrategias de muestreo Si los datos muestrales no se reúnen de forma adecuada, resultarían tan inútiles que ninguna cantidad de tortura estadística podrá salvarlos.

- En una **muestra aleatoria** los miembros de la población se seleccionan de forma que cada miembro individual tenga la misma posibilidad de ser elegido.
- Una **muestra aleatoria simple** de n sujetos se selecciona de manera que cada posible muestra del mismo tamaño n tenga la misma posibilidad de ser elegida.
- Una **muestra probabilística** implica seleccionar miembros de una población de forma que cada miembro tenga una posibilidad conocida (aunque no necesariamente la misma) de ser elegido.

Diseño de experimentos (IV)

3. Aleatoriedad y otras estrategias de muestreo



Muestreo sistemático:
Se selecciona un punto de partida, después se elige cada k -ésimo (por ejemplo, cada quincuagésimo) elemento de la población.



Muestreo de conveniencia:
Se utilizan resultados que son fáciles de obtener.



Muestreo estratificado:
Se subdivide a la población en al menos dos subgrupos (o estratos diferentes), de manera que los sujetos del mismo subgrupo compartan las mismas características (como el género o la categoría de edad), y después se obtiene una muestra de cada subgrupo.



Muestreo por conglomerados:
Se divide el área de la población en secciones (o conglomerados), luego se eligen al azar algunos de estos conglomerados, y después se elige a todos los miembros de los conglomerados seleccionados.

El muestreo por conglomerados usa a todos los miembros de una muestra de conglomerados; en tanto que el muestreo estratificado emplea una muestra de los miembros de todos los estratos.

3. Aleatoriedad y otras estrategias de muestreo

Un **diseño de muestreo de etapas múltiples** implica la selección de una muestra en diferentes pasos, los cuales suelen incluir distintos procedimientos de muestreo.

Nota

Es un diseño de muestreo complejo; no obstante, es mucho más práctico y menos costoso que utilizar un diseño más sencillo, como el muestreo aleatorio simple.

No importa lo bien que usted planee y ejecute el proceso de recolección de muestras, es probable que ocurra algún error en los resultados. Por ejemplo, seleccione 1000 adultos al azar, pregúnteles si se graduaron de bachillerato y registre el porcentaje de respuestas afirmativas en la muestra. Si usted elige otra muestra de 1000 adultos al azar, es probable que obtenga un porcentaje diferente en esa muestra. Esto se conoce como **variación del muestreo**

- Un **error de muestreo** es la diferencia entre el resultado de una muestra y el verdadero resultado de la población; este error es consecuencia de las fluctuaciones por el azar.
- Un **error que no es de muestreo** sucede cuando los datos muestrales se obtienen, registran o analizan de forma incorrecta (como cuando se selecciona una muestra sesgada, cuando se usa un instrumento de medición defectuoso o cuando se copian los datos de forma incorrecta).

Otros Errores comunes (I)

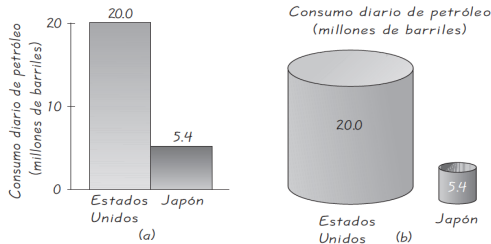
Las conclusiones no se deben basar en muestras demasiado pequeñas.

Gráficas



Otros Errores comunes (II)

Pictogramas



Preguntas predisuestas

- “¿Debería el presidente utilizar su poder de veto para eliminar los desperdicios?” 97 %
- “¿Debería el presidente utilizar su poder de veto o no?” 57 %

Orden de las preguntas

- ¿Cree usted que el tránsito vehicular contribuye a la contaminación del aire más o menos que la industria?
- ¿Cree usted que la industria contribuye a la contaminación del aire más o menos que el tránsito vehicular?

Cuando se presentó primero el tránsito, el 45 % culpó a este factor, y el 27 % a la industria; cuando la industria se presentó primero, el 24 % culpó al tránsito y el 57 % culpó a la industria.

Datos Faltantes

Correlación y causalidad

Correlación no implica causalidad. Esto significa que cuando encontramos una asociación estadística entre dos variables, no podemos concluir que una de las variables sea causa de (o lo que afecta directamente a) la otra variable.

Otros Errores comunes (IV)

Estudios para el propio beneficio Los estudios reciben el patrocinio de grupos con intereses específicos que buscan promover

Números precisos En la actualidad 16,620,021 es el número de habitantes en el Ecuador. En este caso, el número es un estimado y sería mejor decir que el número de habitantes es aproximadamente de 16 millones.

Imágenes parciales “El 90 % de todos nuestros automóviles, vendidos en este país en los últimos 10 años, continúa circulando”. Sin embargo el 90 % de los automóviles que el anunciante vendió en este país se vendieron durante los últimos tres años.

Ejercicios en clase (I)

- 1 **Una ingeniero que supervisa la calidad quiere inspeccionar rollos de papel tapiz para obtener información acerca de la tasa de fallas que tiene la imprenta. Decide tomar una muestra de 50 rollos de la producción de un día. Cada hora durante cinco horas, toma los diez últimos rollos producidos y cuenta el número de fallas de cada uno. ¿Ésta es una muestra aleatoria simple?**
- 2 **Un ingeniero civil acaba de recibir una remesa de mil bloques de hormigón, que pesan aproximadamente 50 libras cada uno. Los bloques se han entregado en una gran pila. El ingeniero quiere investigar la fuerza de compresión de los bloques midiendo las fuerzas en una muestra de diez bloques. Para tomar una muestra aleatoria simple se requeriría sacar bloques del centro y de la parte inferior de la pila, lo que puede ser muy difícil. Por esta razón, el ingeniero puede tomar una muestra simplemente tomando diez bloques de la parte superior de la pila.**

Ejercicios en clase (II)

- 3 Un inspector de calidad prueba 40 pernos de una gran remesa y mide la longitud de cada uno. Descubre que 34 de ellos (85 %) cubre la especificación de longitud. Llega entonces a la conclusión de que exactamente 85 % de los pernos de la remesa satisfacen la especificación. Por otra parte, el supervisor del inspector concluye que la proporción de pernos buenos está cerca de 85 % con cierta probabilidad, pero que no es exactamente igual. ¿Cuál es la conclusión correcta?
- 4 Determine si la descripción dada corresponde a un estudio observacional o a un experimento. Control de calidad. La Food and Drug Administration de Estados Unidos elige al azar una muestra de grageas de aspirina Bayer, y mide la exactitud de la cantidad de aspirina en cada gragea.

Ejercicios en clase (III)

Identifique el tipo de muestreo que se utilizó: aleatorio, sistemático, de conveniencia, estratificado o por conglomerados.

- ⑤ Un investigador de la Universidad de Guayaquil obtiene datos sobre los efectos del alcohol al conducir, examinando informes de accidentes automovilísticos de los últimos cinco años.
- ⑥ ECUAVISA encuesta a 5000 hogares para determinar la proporción de éstos que sintonizan el noticiero de las 20 horas.
- ⑦ Psicología del trauma. Un investigador del hospital Monte Sinaí de la ciudad de Nueva York, planea obtener datos al hacer un seguimiento (hasta el año 2015) a los hermanos de las víctimas que perecieron en el ataque terrorista al Word Trade Center el 11 de septiembre de 2001.
- ⑧ Puesto de revisión de sobriedad donde se detenía y entrevistaba a cada quinto conductor.

Ejercicios en clase (IV)

- 9 Educación y deportes. Un investigador de la empresa de equipo deportivo Spaulding estudia la relación entre el nivel académico y la participación en cualquier deporte. El investigador hace una encuesta a 40 golfistas, 40 tenistas y 40 nadadores, todos elegidos al azar.
- 10 Hacer trampa. Un investigador del Internal Revenue Service estudia las trampas en las declaraciones de impuestos, al encuestar a todos los meseros y las meseras de 20 restaurantes seleccionados al azar.
- 11 Encuestas telefónicas. En una encuesta de Gallup de 1059 adultos, los sujetos encuestados fueron seleccionados, usando de una computadora para generar aleatoriamente los números telefónicos a los que después se llamó.
- 12 Estudiantes que beben. La Universidad de Guayaquil, motivada por un estudiante que murió en estado de ebriedad, realizó una investigación de estudiantes que beben, seleccionando al azar 10 diferentes salones de clase y entrevistando a todos los estudiantes en cada uno de esos grupos.

Ejercicios en clase (V)

- 13 Muestra de conveniencia. Un profesor de estadística obtiene una muestra de estudiantes, al seleccionar a los primeros 10 que entran a su salón de clases. ¿Este plan de muestreo da como resultado una muestra aleatoria? ¿Una muestra aleatoria simple? Explique.
- 14 Muestreo de estudiantes. Un salón de clases tiene 36 estudiantes sentados en seis filas diferentes, con seis estudiantes en cada fila. El profesor tira un dado para determinar una fila, y luego lo tira nuevamente para elegir a un estudiante específico de la fila. Este proceso se repite hasta completar una muestra de 6 estudiantes. ¿Este plan de muestreo da como resultado una muestra aleatoria? ¿Una muestra aleatoria simple? Explique.