

Estadística General

Tema 5: Inferencia Estadística

Generalidades, respuestas aleatorizadas.



Prof. José G. Páez

- Parte de los conceptos de la teoría del muestreo han sido discutidos con anterioridad. Aquí los repasaremos y ampliaremos. Por ejemplo, hemos mencionado que las poblaciones están formadas por individuos, pero sería mejor denominarlas **unidades de muestreo** o **unidades de estudio**:
 - Personas, células, familias, hospitales, países...
- La población ideal que se **pretende estudiar** se denomina **población objetivo**.
 - No es fácil estudiarla por completo. Aproximamos mediante muestras que den idealmente la misma probabilidad a cada individuo de ser elegido.
 - Tampoco es fácil elegir muestras de la población objetivo:
 - Si llamamos por teléfono excluimos a los que no tienen.
 - Si elegimos individuos en la calle, olvidamos los que están trabajando...
- El grupo que en realidad **podemos estudiar** (v.g. los que tienen teléfono) se denomina **población de estudio**.

Fuentes de sesgo

- Las poblaciones objetivo y de estudio **pueden diferir** en cuanto a las variables que estudiamos.
 - El nivel económico en la población de estudio es mayor que en la objetivo,...
 - Los individuos que se eligen en la calle pueden ser de mayor edad (mayor frecuencia de jubilados p.ej.)...
 - En este caso, diremos que las muestras que se elijan estarán **sesgadas**. Al tipo de sesgo debido a diferencias sistemáticas entre población objetivo y población de estudio se denomina **sesgo de selección**.
- Hay otras fuentes de error/sesgo
 - **No respuesta** a encuestas embarazosas
 - Consumo de drogas, violencia doméstica, prácticas poco éticas,...
 - **Mentir** en las preguntas “delicadas”.
- Para evitar este tipo de sesgo se utilizan la técnica de **respuesta aleatorizada**.

Técnicas de respuesta aleatorizada

- Reducen la motivación para mentir (o no responder) a las encuestas.
 - ¿Si digo la verdad, se me verá el plumero...?
- ¿Cómo se hace?
Pídele que lance una moneda antes de responder y...
 - Si sale **cara** que diga la “**opción comprometida**”
 - (no tiene por qué avergonzarse, la culpa es de la moneda)
 - Si sale **cruz** que diga la **verdad**
 - (no tiene por qué avergonzarse, el encuestador no sabe si ha salido cara o cruz)
- Aunque no podamos saber cuál es la verdad en cada individuo, podemos hacernos una idea porcentual sobre la población, viendo en cuánto se alejan las respuestas del 50%.

Ejemplo: ¿Ha tomado drogas alguna vez?

Sin respuesta aleatorizada

100% No



Insineros!!



Con respuesta aleatorizada

40% No
60% Sí

Diferencia entre los que han dicho sí y los que debían hacerlo por que así lo indicaba la moneda

$$p^* \approx \frac{0,6 - 0,5}{1 - 0,5} = 0,2 = 20\%$$

¡No son mitad y mitad!

El porcentaje estimado de individuos que tomó drogas es: 20%

Los que deben decir la verdad