

## Trabajo en clase. Recolección y Muestreo de Datos

### 1. Población, muestra, parámetros y estadísticos

- Los datos son información proveniente de observaciones, conteos, medidas o respuestas.
- La estadística es la ciencia (y el arte) de diseñar estudios, así como de analizar e interpretar los datos que esos estudios producen.
- Una **población** es el conjunto de todos los sujetos que nos interesan.
- Una **muestra** es un subconjunto de una población. Este es el grupo del que tenemos datos.
- Un **parámetro** es una descripción numérica de una característica de la **población**.
- Un **estadístico** es una descripción numérica de una característica de la **muestra**.
- Las estadísticas descriptivas se refieren a métodos para resumir los datos recopilados. Estos resúmenes generalmente implican hacer gráficos, encontrar promedios y porcentajes.
- La estadística inferencial involucra el uso de datos de una muestra para sacar conclusiones sobre una población.

### Preguntas 1

Para cada uno de los siguientes escenarios, identifique la población (implícita), la muestra y cualquier parámetro y estadístico:

- a) En California, en octubre de 2003, se llevó a cabo una elección especial para determinar si el Gobernador debería ser destituido de su cargo. La encuesta de salida en la que se basaron las cadenas de televisión para sus proyecciones encontró (después de muestrear a 3160 votantes) que el 54% de la muestra votó para destituir al Gobernador y el 46% votó para no destituirlo.
- b) Un noticiero de televisión informó que el 44,9% de los encuestados (1261 adultos) comía en restaurantes de comida rápida de una a tres veces por semana.
- c) Según un estudio realizado en Estados Unidos por la empresa Consumer Reports, una muestra aleatoria de 35 autos nuevos reportó un promedio de rendimiento de gasolina de 21.1 millas por galón.
- d) 4 de un total de 524 pasajeros y tripulantes sobrevivieron al accidente del vuelo 123 de Japan Airlines en agosto de 1985.

## 2. Recopilación de datos

Nos centraremos en cuatro métodos para recopilar datos: estudios observacionales, experimentos, simulaciones, y encuestas.

- En un estudio observacional, un investigador mide y observa las variables de interés sin cambiar las condiciones existentes.
- En un experimento, un investigador asigna un tratamiento y observa la respuesta. A veces, se puede usar un grupo de control (un grupo que no recibe tratamiento o recibe un placebo ) para comparar la efectividad de un tratamiento.
- Una simulación utiliza un modelo matemático, físico o informático para replicar las condiciones de un proceso o situación. Esto se usa con frecuencia cuando la situación real es demasiado costosa, peligrosa o poco práctica para reproducirla en la vida real.
- Una encuesta se utiliza para investigar las características de una población. Se usa con frecuencia cuando los sujetos son personas y se les hacen preguntas.

## Preguntas 2

Identifique qué método para recopilar datos (estudio observacional, experimento, simulación o encuesta) es mejor en cada una de las siguientes situaciones y explique su respuesta.

- a) El efecto que tendría un fuerte terremoto en Cartago.
- b) Si un cupón de descuento adjunto a un correo electrónico de publicidad hace que los receptores de la publicidad sean más propensos a pedir productos de una empresa.
- c) Si fumar tiene o no un efecto sobre la enfermedad coronaria.
- d) Determinar el ingreso familiar promedio de los hogares en San isidro de Heredia.

## 3. Técnicas de Muestreo

- Un censo es utilizar como sujeto a cada miembro de una población. En caso de contar con un censo, las estadísticas descriptivas serían suficientes. Sin embargo, esto es a menudo demasiado costoso y difícil. En su lugar, muestreamos parte de la población. Con el muestreo, debemos asegurarnos de que la muestra sea representativa de la población y lo suficientemente grande como para ser significativa.
- Un error de muestreo es la diferencia entre los resultados de la muestra y los de la población. Incluso con las mejores técnicas de muestreo, esto es posible.
- Una muestra sesgada es aquella que no es representativa de toda la población. Queremos evitar sesgos.

- Una muestra aleatoria es aquella en la que todos los miembros de la población tienen las mismas posibilidades de ser elegidos.
- Una muestra aleatoria simple (SRS) es una muestra en la que todas las muestras posibles del mismo tamaño tienen la misma probabilidad de ser recolectadas. Normalmente, comenzaremos usando una muestra aleatoria simple.
- Se usa una muestra estratificada cuando es importante tener miembros de múltiples segmentos de la población. Primero, la población se divide en segmentos (llamados "estratos"), luego se elige un número predeterminado de sujetos de cada uno de los estratos.
- El muestreo por conglomerados se puede utilizar cuando la población cae naturalmente en subgrupos con características similares. Primero, determine los conglomerados, luego seleccione todos los miembros de uno o más de los conglomerados (o un subconjunto usando una muestra aleatoria).
- El muestreo sistemático implica primero asignar un número a cada miembro de la población y ordenarlos de alguna manera. Los miembros de la muestra se seleccionan eligiendo el primer miembro al azar y luego seleccionando los miembros subsiguientes a intervalos regulares después del número inicial (por ejemplo, cada 7 personas). Este método es simple de usar, pero debe evitarse si hay patrones que ocurren regularmente en los datos.
- Una muestra de conveniencia consta solo de miembros disponibles de la población, pero esto a menudo conduce a estudios sesgados.
- Una muestra voluntaria es una especie de muestra de conveniencia en la que solo participan voluntarios.

### Preguntas 3

Determine qué tipo de muestreo se utilizó en cada uno de los siguientes escenarios (aleatorio simple, estratificado, por conglomerados, sistemático, de conveniencia, voluntaria, censo, ...):

- a) Para determinar la calidad de los dormitorios en el campus universitario de la Universidad de Marsella, se eligieron 20 residentes de cada uno de los 3 dormitorios y se les pasó una encuesta. Hay dos edificios de dormitorios con servicios compartidos para los estudiantes de pregrado (uno para hombres, otro para mujeres) y un conjunto de dormitorios individuales (con baño propio) para estudiantes de posgrado.
- b) Para evaluar la compensación de los empleados de una empresa con 6000 empleados, se eligieron 4 departamentos al azar (resultando elegidos Recursos Humanos, Mantenimiento Eléctrico, Oficiales de Seguridad y Contabilidad), luego se encuestaron a todas las personas dentro de cada uno de esos departamentos.
- c) Un grupo de personas que participan en una encuesta que se publicó en el sitio cnn.com invitando a los visitantes del sitio a emitir su opinión sobre ChatGPT.

- d) Para determinar la percepción de la calidad de la educación en la UCR, se van a seleccionar 30 estudiantes que ingresaron en el año 2020 (carné C0xxxx). Para establecer la parte variable de los números de carné (xxxx) se elige un número “z” entre 0 y 9 al azar y luego se selecciona un estudiante cada 250 a partir de ese z:  $z+250$ ,  $z+(250*2)$ ,  $z+(250*3)$ , ... ,  $z+(250*3)$ .
- e) Interesado en la seguridad de su cantón, usted decide hacer una encuesta en su barrio, yendo de puerta en puerta para hacer preguntas a los vecinos. Dado que su tiempo es limitado, no tiene la oportunidad de volver a visitar casas donde nadie respondió a la puerta.
- f) Interesado en la seguridad de su barrio, usted va de puerta en puerta para hacer preguntas a los vecinos. Todos estaban en casa y dispuestos a participar, por lo que tiene los resultados de la encuesta de cada hogar en el barrio.
- g) Escogidos al azar de entre personas que recibieron atención en el Centro de Salud de la Universidad durante el último año, 300 personas participaron en una encuesta sobre la atención recibida.