Introducción al Muestreo y Diseño experimental

Aniel Nieves-González

Instituto de estadísticas

¿Cómo estudiar a los individuos (o sujeto) de interés?

• Estudio observacional: observa individuos, mide variable de interés pero no intenta influenciar (perturbar) la respuesta de estos.

¿Cómo estudiar a los individuos (o sujeto) de interés?

- Estudio observacional: observa individuos, mide variable de interés pero no intenta influenciar (perturbar) la respuesta de estos.
- Experimento define tratamiento (perturbación) sobre los individuos y observa su respuesta.

¿Cómo estudiar a los individuos (o sujeto) de interés?

- Estudio observacional: observa individuos, mide variable de interés pero no intenta influenciar (perturbar) la respuesta de estos.
- Experimento define tratamiento (perturbación) sobre los individuos y observa su respuesta.

¿Cómo estudiar a los individuos (o sujeto) de interés?

- Estudio observacional: observa individuos, mide variable de interés pero no intenta influenciar (perturbar) la respuesta de estos.
- Experimento define tratamiento (perturbación) sobre los individuos y observa su respuesta.

Una idea fundamental en ambos acercamientos, es la de **muestreo** (sampling).

Definition (Población (Population))

Conjunto de individuos (u objetos) sobre el cual queremos información.

Definition (Población (Population))

Conjunto de individuos (u objetos) sobre el cual queremos información.

Definition (Muestra (Sample))

Subconjuto de la población sobre el cual obtenemos información. Con este, inferimos conclusiones sobre la población.

Definition (Población (Population))

Conjunto de individuos (u objetos) sobre el cual queremos información.

Definition (Muestra (Sample))

Subconjuto de la población sobre el cual obtenemos información. Con este, inferimos conclusiones sobre la población.

Es de crucial importancia definir *apropiadamente* ambos conjuntos al momento de empremder algún estudio.

- Encuestas.
- Censos.
- Sacar sangre u obtener orina para algún test médico.

Decimos que el diseño de un estudio (o la selección de una muestra) está sesgada (bias), si sistemáticamente se favorecen ciertos resultados

Por ejemplo:

• "Voluntary response sample". En este caso un subconjunto de la población piden ser los representantes de la misma (e.g textear a "American Idol").

Decimos que el diseño de un estudio (o la selección de una muestra) está sesgada (bias), si sistemáticamente se favorecen ciertos resultados

- "Voluntary response sample". En este caso un subconjunto de la población piden ser los representantes de la misma (e.g textear a "American Idol").
- Potencialmente llamar por télefono para encuestar.

Decimos que el diseño de un estudio (o la selección de una muestra) está sesgada (bias), si sistemáticamente se favorecen ciertos resultados

- "Voluntary response sample". En este caso un subconjunto de la población piden ser los representantes de la misma (e.g textear a "American Idol").
- Potencialmente llamar por télefono para encuestar.
- Caso elecciones EEUU: Encuestar en las inmediaciones (temporal o físicas) de la convención de alguno de los partidos principales.

Decimos que el diseño de un estudio (o la selección de una muestra) está sesgada (bias), si sistemáticamente se favorecen ciertos resultados

- "Voluntary response sample". En este caso un subconjunto de la población piden ser los representantes de la misma (e.g textear a "American Idol").
- Potencialmente llamar por télefono para encuestar.
- Caso elecciones EEUU: Encuestar en las inmediaciones (temporal o físicas) de la convención de alguno de los partidos principales.
- Tomar *una* sola muestra de sangre para determinar nivel de glucosa en la sangre y así determinar función pancreatica.

Decimos que el diseño de un estudio (o la selección de una muestra) está sesgada (bias), si sistemáticamente se favorecen ciertos resultados

- "Voluntary response sample". En este caso un subconjunto de la población piden ser los representantes de la misma (e.g textear a "American Idol").
- Potencialmente llamar por télefono para encuestar.
- Caso elecciones EEUU: Encuestar en las inmediaciones (temporal o físicas) de la convención de alguno de los partidos principales.
- Tomar *una* sola muestra de sangre para determinar nivel de glucosa en la sangre y así determinar función pancreatica.
- Lea artículo: The art of election polling. SIAM News (Nov. 26 2004), por Sara Robinson. (Hay link en la página de la clase, bajo referencias).

Muestreo (sampling)

Definition (Probability sample)

Escoger la muestra aleatoriamente. Para esto debemos conocer las probabilidades (chances) de cada posible muestra.

Existen varios tipos de muestreo probabilistico.

Muestreo (sampling)

Definition (Simple Random Sample SRS (muestreo aleatorio simple))

Muestra de n individuos escogidos de la población de manera en que cada subconjunto de n individuos tiene el mismo chance (probabilidad) de ser escogido.

Este tipo de muestreo es fundamental en terminos prácticos y teóricos.

- Para escoger un SRS se puede hacer uso de números pseudoaleatorios (PRN, en inglés).
- Los PRN son una sucesión de núm. generados por una computadora (algorítmicamente) que aproximan una sucesión de números aleatorios.

- Son PRN, esto es:
 - ► Son cíclicos...

- Son PRN, esto es:
 - ► Son cíclicos...
 - ▶ Cada sucesión depende de una semilla...

- Son PRN, esto es:
 - ► Son cíclicos...
 - Cada sucesión depende de una semilla...
- Tienen propiedades estadisticas específicas (siguen alguna distribución).

- Son PRN, esto es:
 - ► Son cíclicos...
 - Cada sucesión depende de una semilla...
- Tienen propiedades estadisticas específicas (siguen alguna distribución).
- Son imp. en diferentes aplicaciones: Criptografía, Metodos Monte Carlo, etc.

 $\ensuremath{\mathsf{\mathcal{C}}}$ Como escoger un SRS?

Algorithm

- Asigne un "label" (etiqueta) a cada miembro de la población. La etiqueta bien puede ser un número natural.
- 2 Use una generador de números pseudoaleatorios o una tabla con números pseudoaleatorios precalculados para escoger muestra.

Mire ejemplo 3.8. Para que vea como se usa una rabla de números pseudoaleatorios. Mire script en R mostrado en clase.

Muestreo (sampling)

Definition (Muestra aleatoria estratificada (Stratified Random Sample))

Muestra para la cual se divide la población en grupos de individuos semejantes (estratas). Se escoge un SRS de cada estrato y se combina para construir una muestra de la población.

Note que la clave es dividir la población en grupos con características particulares.

- Existen otros tipos de muestreo.
- Podemos enfrentar dificultades con un muestreo si existe:
 - undercoverage: ocurre cuando algún grupo de la población queda fuera del muestreo (¿Estamos muestrando la población?).

- Existen otros tipos de muestreo.
- Podemos enfrentar dificultades con un muestreo si existe:
 - undercoverage: ocurre cuando algún grupo de la población queda fuera del muestreo (¿Estamos muestrando la población?).
 - ▶ nonresponse: ocurre cuando un individuo decide no ser parte de la muestra. (Crítico en ciencias sociales, encuestas, etc.)

Diseño experimental

Recordemos que:

• Experimento define tratamiento (perturbación) sobre los individuos y observa su respuesta.

Mucha de la terminología que estudiaremos es la usada en el mundo de los clinical trials.

Terminología

- Los individuos son llamados sujetos.
- La variable explicativa (independiente) es llamada factor.
- Tratamiento es la condición experimental impuesta a los sujetos.
- En general un experimento puede esquematisarse así:

 $sujetos \rightarrow tratamiento \rightarrow respuesta$

Suponga que a una farmaceutica le interesa vender un medicamento para el sickle-cell disease (hydroxyurea). El FDA permite la salida al mercado de nuevos medicamentos luego de la realización de experimentos (clinical trials). El clinical trial dió la hydroxyurea a 150 pacientes y un placebo a otros 150. Los investigadores contaron el número de episodios reportados por los pacientes de ciertos síntomas de la enfermedad. ¿Cuáles son los sujetos, el tratamiento, el response variable, y los factores?

¿Cuáles son los sujetos, el tratamiento, el response variable y el factor?

• Sujetos: los 300 participantes del estudio.

¿Cuáles son los sujetos, el tratamiento, el response variable y el factor?

- Sujetos: los 300 participantes del estudio.
- Tratamiento: dar la hydroxyurea versus placebo.

¿Cuáles son los sujetos, el tratamiento, el response variable y el factor?

- Sujetos: los 300 participantes del estudio.
- Tratamiento: dar la hydroxyurea versus placebo.
- Response variable (variable dependiente): número de episodios de el síntoma (dolor).

¿Cuáles son los sujetos, el tratamiento, el response variable y el factor?

- Sujetos: los 300 participantes del estudio.
- Tratamiento: dar la hydroxyurea versus placebo.
- Response variable (variable dependiente): número de episodios de el síntoma (dolor).
- Factor: tipo de medicamento (hydroxyurea vs. placebo).

• Es importante controlar o aislar al experimento del ambiente donde se realiza para evitar confundir factores.

- Es importante controlar o aislar al experimento del ambiente donde se realiza para evitar confundir factores.
- ② En un experimento comparativo se comparan tratamientos.

- Es importante controlar o aislar al experimento del ambiente donde se realiza para evitar confundir factores.
- ② En un experimento comparativo se comparan tratamientos.
 - ▶ Se compara a un grupo bajo tratamiento y a otro que no.

- Es importante controlar o aislar al experimento del ambiente donde se realiza para evitar confundir factores.
- ② En un experimento comparativo se comparan tratamientos.
 - ▶ Se compara a un grupo bajo tratamiento y a otro que no.
 - ▶ El grupo que no está bajo tratamiento se le llama control.

- Es importante controlar o aislar al experimento del ambiente donde se realiza para evitar confundir factores.
- 2 En un experimento comparativo se comparan tratamientos.
 - ▶ Se compara a un grupo bajo tratamiento y a otro que no.
 - ▶ El grupo que no está bajo tratamiento se le llama control.
 - ► Es importante asignar a los sujetos a distintos tratamientos en forma aleatoria (randomization).

- Es importante controlar o aislar al experimento del ambiente donde se realiza para evitar confundir factores.
- ② En un experimento comparativo se comparan tratamientos.
 - ▶ Se compara a un grupo bajo tratamiento y a otro que no.
 - ▶ El grupo que no está bajo tratamiento se le llama control.
 - ► Es importante asignar a los sujetos a distintos tratamientos en forma aleatoria (randomization).
 - ▶ El randomization minimiza sesgos del experimentalista y cargar a un tratamiento con individuos similares.

- Es importante controlar o aislar al experimento del ambiente donde se realiza para evitar confundir factores.
- ② En un experimento comparativo se comparan tratamientos.
 - ▶ Se compara a un grupo bajo tratamiento y a otro que no.
 - ▶ El grupo que no está bajo tratamiento se le llama control.
 - ► Es importante asignar a los sujetos a distintos tratamientos en forma aleatoria (randomization).
 - ▶ El randomization minimiza sesgos del experimentalista y cargar a un tratamiento con individuos similares.
 - ▶ Decimos que el experimento es de tipo double-blind cuando ni el experimentalista sabe q. individuos están en cual tratamiento.

- Es importante controlar o aislar al experimento del ambiente donde se realiza para evitar confundir factores.
- 2 En un experimento comparativo se comparan tratamientos.
 - ▶ Se compara a un grupo bajo tratamiento y a otro que no.
 - ▶ El grupo que no está bajo tratamiento se le llama control.
 - ► Es importante asignar a los sujetos a distintos tratamientos en forma aleatoria (randomization).
 - ► El randomization minimiza sesgos del experimentalista y cargar a un tratamiento con individuos similares.
 - ▶ Decimos que el experimento es de tipo double-blind cuando ni el experimentalista sabe q. individuos están en cual tratamiento.
- Ompletely randomized experiment: se asignan sujetos de forma aleatoria a distintos tratamientos.

Definition (Statistical significance (estadisticamente significativo))

Decimos que un efecto observado es estadísticamente significativo si es de tal magnitud que sería *muy raro* que ocurriera aleatoriamente.

Estudie lo siguiente:

- Concepto de block design: se agrupan individuos que se sabe que son similares. La aleatorización de las asignaciones a tratamientos se hace separadamente.
- Concepto de matched pairs: caso especial del block design.

Resumimos los principios de diseño experimental:

- Se **controlan** los efectos de variables asomantes (lurking variables) mediante la comparación de distintos tratamientos (incluye grupo control).
- Aleatorización de la asig. de individuos a tratamientos.
- Repite los tratamientos en un núm. de individuos suficientemente grande para minimizar variabilidad.

 ${\bf Preguntas...}$