

Medición

Fill In Your Name

02 January, 2022

Una parte clave de su diseño de investigación

Conceptos y mediciones

De los conceptos a los datos o puntuaciones I

De los conceptos a los datos o puntuaciones II

Qué se debe medir

Se debería medir todo

Los indicadores deben conectarse con las teorías y los mecanismos

Cómo hacer mediciones

Error de medición y sus consecuencias

Limitación del error de medición

Una parte clave de su diseño de investigación

Medición

- ▶ La medición es una parte esencial del diseño de la investigación.
- ▶ La medición se desprende de su teoría sobre la forma en usted cree que funciona el mundo, y cómo cree que su tratamiento manipula ese mundo. Es decir, la medición es un ejercicio tanto teórico como técnico y logístico.
- ▶ Cuando registramos un número, una palabra o una letra en un conjunto de datos para **representar** algo abstracto en el mundo (como “estabilidad financiera”, “hambre” o “capacidad matemática”), estamos **midiendo**.

Una mala medición es problemática

- ▶ Los problemas de medición pueden llevarle a hacer inferencias (causales) incorrectas del estudio (error sistemático).
- ▶ Las mediciones ruidosas reducen el poder (error aleatorio).

Recogida de datos

- ▶ La recopilación de datos suele ocupar una gran parte del tiempo y de los recursos financieros disponibles en el presupuesto del proyecto.
- ▶ Los datos nuevos pueden ser un resultado útil de la investigación por sí mismos y una base importante para futuras investigaciones. ¡Los datos son un bien público!

Conceptos y mediciones

De los conceptos a los datos o puntuaciones I

De los conceptos a los datos o puntuaciones II

De los conceptos a los datos o puntuaciones II

- ▶ ¿Cuál es un ejemplo de un concepto (como una variable de resultado) que le gustaría cambiar con un tratamiento experimental?
- ▶ ¿Cómo podría saber que una unidad (como una persona o un pueblo) muestra una “capacidad” o un “rango” alto o bajo sobre ese concepto? (¿Era “más hambrienta” o “más pacífica”?)
- ▶ ¿Qué podría observar para sentirse más o menos seguro de que una persona en particular era más hambrienta o menos partidaria de la violencia que otra persona?
- ▶ ¿Qué críticas podría recibir si dijera: “Creo que esta unidad difiere de esta otra unidad en mi variable de resultado conceptual”?

Validez de la medición

- ▶ Aquí hablamos de “**validez** de la medición”.
- ▶ Una medida válida es aquella que podemos argumentar de forma persuasiva (que verdaderamente representa lo que decimos que representa, y que no representa otros aspectos de la unidad.) Véase Shadish et al. (2002) para más información sobre la validez y la fiabilidad de la medición).

Pregunta I

- ▶ *“Un barco que puede recorrer cuarenta millas por hora en aguas tranquilas hace un viaje de cien millas por un determinado arroyo. Si este viaje dura dos horas, ¿cuánto durará el viaje de vuelta?” (del SAT de 1926 en los EE.UU)*
- ▶ ¿Mide esto la “capacidad matemática”? ¿“capacidad verbal”? o ¿el conocimiento sobre navegación?

Pregunta II

Elija el par de palabras más parecidas al par superior:

Corredor : maratón

- (a) enviado : embajada
- (b) mártir : masacre
- (c) remero : regata
- (d) caballo : establo

La respuesta correcta es: (c) remero : regata. (de la prueba SAT de los años 80 en EE.UU.)

- ¿Mide esto el conocimiento de los deportes universitarios de élite? ¿O la “capacidad verbal”?

Qué se debe medir

Se debería medir todo

Se debería medir todo

- ▶ Variables de resultado de todas las unidades de su estudio (para cada ola, y cualquier deserción).
- ▶ Tratamiento (asignación y cumplimiento).
- ▶ Indicaciones de que su tratamiento se recibe e interpreta como usted espera (comprobación de la manipulación).
- ▶ Indicios de que su investigación pueda causar daños.
- ▶ Covariables, incluido el contexto que podría (a) afectar al funcionamiento de su intervención o (b) influir en la variabilidad de su variable de resultado.

Los indicadores deben conectarse con las teorías
y los mecanismos

Los indicadores deben conectarse con las teorías y los mecanismos

- ▶ A menudo tenemos múltiples teorías sobre *cómo* una intervención podría afectar a una variable de resultado (diferentes mecanismos).
- ▶ Medir indicadores que sean *únicos* para cada mecanismo y que puedan ayudar a *diferenciar* entre ellos.
- ▶ Dichos indicadores pueden incluir variables de resultado intermedias que se realizan *antes* de la variable de resultado final.
- ▶ Pueden incluir variables de resultado secundarias, como
 - ▶ variables de resultado para los que esperamos efectos sólo bajo ciertas teorías.
 - ▶ variables de resultado placebo para las que no esperamos efectos.

Indicadores

- ▶ El caso ideal es la medición directa del concepto o fenómeno de interés Sin ningún error (raramente posible).
 - ▶ La asignación del tratamiento bajo su control es la rara excepción y es crucial registrarla.
- ▶ A menudo sólo somos capaces de medir **indicadores** que están relacionados con el concepto o fenómeno de interés, pero no que no son totalmente determinantes en él.
 - ▶ Las respuestas correctas a problemas particulares (indicadores) para la aptitud matemática subyacente (el fenómeno real).
 - ▶ Días sin comer (indicadores) para el hambre (el fenómeno real).

Seleccione indicadores válidos

- ▶ Las personas razonables pueden no estar de acuerdo con la conceptualización. ¿Qué entendemos realmente por aptitud matemática o por hambre? ¿Qué más pueden decirnos nuestras puntuaciones sobre nuestra gente, pueblos, etc. además de lo que esperamos representar?
- ▶ Seleccione los indicadores que se relacionan estrechamente con lo que **queremos decir** cuando hablamos sobre el fenómeno de interés.
- ▶ Seleccionar indicadores **válidos**. Ej. Las mediciones de matemáticas no deberían medir también los conocimientos de navegación.

Selección indicadores fiables

- ▶ Seleccione indicadores que sean **fiables**.
- ▶ Ejemplo: Una vara de medición debe medir siempre un metro sin importar la temperatura. Una varilla de goma para medición puede producir mediciones **poco fiables** del concepto abstracto de longitud, especialmente si lo utiliza un niño de 5 años. Un láser puede producir mediciones más fiables incluso que un palo de madera en manos de un usuario experto en palos de madera.

Medidas múltiples

- ▶ Si tiene *múltiples indicadores para el mismo fenómeno*, tendrá que determinar cómo agregar estos indicadores.
 - Promediar no siempre tendrá sentido. ¿Se le ocurren
- ▶ Disponer de múltiples indicadores a menudo puede mejorar la **fiabilidad** de la medición y quizás logre reforzar los argumentos de **validez**.
- ▶ Combinar el uso de un palo de madera y un láser podría ser ideal para convencernos de que hemos medido la longitud de una manera que entendemos como **longitud** (en lugar de, por ejemplo, la temperatura) y que la puntuación no depende crucialmente del contexto de la medición.

Cómo hacer mediciones

Herramientas de medición y fuentes de datos

- ▶ Después de determinar qué concepto debemos medir y cómo podemos reconocerlo cuando lo vemos, tendremos que pensar en **cómo** medirlo.
- ▶ Disponemos de varias herramientas y fuentes:
 - ▶ Medidas de encuesta
 - ▶ Medidas de comportamiento
 - ▶ Datos administrativos (registros fiscales, resultados electorales, etc.)
 - ▶ Imágenes/detección remota
 - ▶ Texto (transcripciones de voz, periódicos, etc.)
 - ▶ Sensores en vestuario, teléfonos, otros dispositivos (artículos domésticos)
 - ▶ Otros

Consideraciones a la hora de elegir herramientas y fuentes

Diferentes herramientas implican diferentes compensaciones:

- ▶ Validez (¿La puntuación capta el concepto y sólo el concepto?)
 - ▶ Fiabilidad (¿Un procedimiento de puntuación utilizado dos veces daría la misma respuesta?)
 - ▶ Sesgo (error sistemático)
 - ▶ Precisión (error aleatorio)
 - ▶ Muestra para la que se pueden realizar mediciones
 - ▶ Momento de las mediciones
 - ▶ Coste

Mantener la simetría entre los brazos de tratamiento

- ▶ Se debe mantener la simetría entre los grupos de tratamiento y de control mientras se realiza la medición.
- ▶ Muestra para la que se pueden hacer mediciones
 - ▶ El número de veces y la duración de la interacción con los participantes debe ser igual en todos los brazos de tratamiento.
 - ▶ Las preguntas deben ser las mismas.
 - ▶ Tenga especial cuidado con estos puntos para los indicadores utilizados en los controles de manipulación.
- ▶ **La medición no debe ser la intervención experimental.**
Queremos que la única diferencia entre los brazos de tratamiento sea la intervención, no la medición.

Error de medición y sus consecuencias

Error de medición I

- ▶ Queremos evitar/minimizar dos tipos de error de medición:
 - ▶ El error sistemático (sesgo)
 - ▶ Error aleatorio (falta de precisión)

Error de medición II

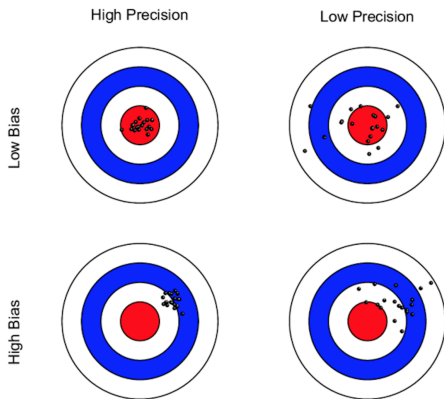


Figure 2: Sesgo y Precisión

(Fuente: Precision and Bias of Masurement)

Ejemplos de errores de medición

- ▶ Error sistemático
 - ▶ Una báscula mal calibrada, por lo que todo el mundo parece pesar 2 kg menos de lo que realmente pesa.
 - ▶ Un diario de alimentación que deja de reportar sistemáticamente los tentempiés.
 - ▶ Efectos de la demanda.
 - ▶ Efectos Hawthorne.
- ▶ Error aleatorio
 - ▶ Una mano temblorosa al marcar la distancia.
- ▶ Tenga en cuenta que estos son ejemplos de mediciones **poco fiables**. También podrían ser **inválidos**, pero no tienen por qué serlo.

Consecuencias de medir mal el tratamiento

- ▶ Además de generar una descripción incorrecta del nivel de algún fenómeno, el error de medición puede afectar nuestras inferencias causales.
- ▶ Si la variable de tratamiento es binaria (una unidad puede estar sólo en el tratamiento o en el control), el error de medición está correlacionado negativamente con la variable real. (Un 1 se codifica erróneamente como 0, por lo que el error es -1 ; un 0 se codifica erróneamente como 1, por lo que el error es 1).
- ▶ Si utiliza OLS para calcular los estimadores de los efectos promedios del tratamiento, este tipo de error conduce a estimaciones más pequeñas del efecto causal (coeficiente sobre la variable de tratamiento).

Consecuencias del error de medición aleatorio

- ▶ En un análisis OLS, un mayor error aleatorio en la variable de resultado conduce a estimaciones menos precisas del efecto causal (coeficiente en la variable de tratamiento).
- ▶ La reducción del error de medición aleatorio en la variable de resultado puede aumentar el poder estadístico (porque la variable de resultado tiene menos ruido que no se relaciona con el tratamiento).

Consecuencias del error de medición sistemático I

- ▶ Si todas las mediciones están desviadas por la misma cantidad, como -2kg para todos
 - ▶ esto no supone ninguna diferencia cuando los efectos del tratamiento se definen como diferencias en las variables de resultado potenciales,

$$\tau_i = Y_i(1) - Y_i(0), \tilde{\tau}_i = (Y_i(1) - 2) - (Y_i(0) - 2), \text{ para que } \tau_i = \tilde{\tau}_i$$

- ▶ pero esto es problemático si el efecto del tratamiento se define como la proporción de variables de resultado ($Y_i(1) > 0$ y $Y_i(0) > 0$).

$$\tau_i = Y_i(1)/Y_i(0), \tilde{\tau}_i = (Y_i(1) - 2)/(Y_i(0) - 2), \text{ para que } \tau_i \neq \tilde{\tau}_i \text{ excepto cuando } Y_i(1) = Y_i(0).$$

Lo lejos que esté $\tilde{\tau}_i$ de τ_i depende de lo grande que sea 2 (el error) en relación con los valores reales $Y_i(0)$, $Y_i(1)$.

Consecuencias del error de medición sistemático II

- ▶ Nota: los coeficientes de regresión logística son cocientes de variables de resultado potenciales.
- ▶ El error de medición puede estar correlacionado con el valor verdadero de Y .
- ▶ Por ejemplo, las personas que se involucran en un comportamiento mal visto, vergonzoso o ilegal pueden reportar menos dicho comportamiento, mientras que los que no lo hacen pueden informar de su nivel con precisión. (Esto se conoce como sesgo de deseabilidad social).
- ▶ Este fenómeno puede ocurrir con las víctimas de este tipo de comportamiento por parte de otras personas, que se culpan a sí mismas o temen ser sancionadas por otras personas que preferirían no conocer el comportamiento (por ejemplo, las víctimas de la violencia de pareja).

Consecuencias del error de medición sistemático III

- ▶ Esto hace que sea más difícil detectar el efecto de una intervención diseñada para reducir este comportamiento.
- ▶ Otra forma de sesgo de deseabilidad social también puede hacer que el error de medición esté correlacionado con el tratamiento.
- ▶ Por ejemplo, su intervención podría tener como objetivo reducir las actitudes hostiles hacia los miembros de otros grupos sociales. Si los participantes pueden averiguar los objetivos de su estudio, es posible que (inconscientemente) traten de complacer al investigador diciéndole lo que quiere ver. Los del grupo de tratamiento pueden reportar en menor medida su hostilidad hacia otros grupos en comparación con el grupo de control.

Consecuencias del error de medición sistemático IV

- ▶ Esto hace que sea difícil saber si la diferencia en las variables de resultado observadas entre los grupos de tratamiento y de control se debe a que la intervención realmente reduce la hostilidad o a que el conocimiento del tratamiento cambia el reporte de la hostilidad.

Limitación del error de medición

Algunas opciones para limitar los errores de medición I

- ▶ El auto reporte de un sujeto (en una encuesta) es más problemático que la observación discreta del sujeto (“en la naturaleza”) por otra persona.
- ▶ Las medidas de comportamiento están menos sujetas al sesgo de deseabilidad social.
- ▶ Los registros administrativos para los que los informes erróneos tienen sanciones legales podrían ser más precisos.
- ▶ Proporcionar más privacidad para que la puntuación pueda producirse sin la observación de otras personas o del investigador.
- ▶ Mantener algunas hipótesis y objetivos del estudio ocultos a los participantes.

Algunas opciones para limitar los errores de medición II

- ▶ Si no se puede controlar el error de medición, hay que estudiarlo para saber si es un problema y que tan grande es su magnitud. Considerar los estudios piloto orientados a la medición.

Ejemplo: registros administrativos

- ▶ Registros de asistencia a una reunión, en lugar de preguntar si alguien asistió.
 - ▶ Es posible que sólo existan registros de asistencia regulares para las reuniones que no son de interés para su población objetivo original.
 - ▶ Es posible que tenga que planificar con antelación la recogida de datos en una reunión futura.

Ejemplo - medidas de comportamiento I

- ▶ Pida a los sujetos que realicen un esfuerzo inducido, como firmar una petición, hacer una donación o realizar alguna otra tarea que tenga un pequeño coste personal, en lugar de preguntarles si apoyan un tema concreto.
 - ▶ Esto puede captar sólo a los sujetos que se preocupan bastante por ese tema.
 - ▶ Ejemplo: El trabajo de Pedro Vicente sobre la compra de votos en Sao Tomé y Príncipe utiliza encuestas, registros administrativos y si los encuestados enviaron una tarjeta postal pre-pagada para medir las variables de resultado. Léase [Brief 20: Is Vote Buying Effective?](#)

Ejemplo - medidas de comportamiento II

- ▶ Realizar “juegos de laboratorio” para medir la cooperación o la generosidad hacia los grupos externos, en lugar de preguntar a los sujetos si cooperarían con otros.
 - ▶ El estudio de Scacco y Warren sobre el prejuicio y la discriminación utiliza variantes de juegos de dictador. Léase [Can Social Contact Reduce Prejudice and Discrimination? Evidence from a Field Experiment in Nigeria](#)

Ejemplo - anonimidad o privacidad I

- ▶ Asegurar “anonimidad” para que los encuestados crean que sus respuestas no pueden ser rastreadas a ellos mismos.
 - ▶ Respuesta aleatoria: el azar determina si el encuestado debe responder a una pregunta con la verdad o responder de manera afirmativa, independientemente de cuál sea la verdad. El encuestador no sabe en qué condición puso el azar al encuestado.

Ejemplo: anonimidad o privacidad II

- ▶ Experimentos con listas: se da a los encuestados una lista de elementos o afirmaciones y se les pregunta cuántos son verdaderos para ellos. Los encuestados se reparten aleatoriamente entre diferentes listas, una de las cuales contiene un elemento sensible adicional (por ejemplo, “mi marido me pega”). Esto permite al investigador estimar la prevalencia de un elemento concreto. Tenga en cuenta que este enfoque reduce el poder para un tamaño de muestra determinado.
- ▶ Privacidad sencilla: en el caso de preguntas como la elección del voto, pida al encuestado que rellene una papeleta falsa y la deposite en una caja cerrada en lugar de responder directamente al encuestador.

Ejemplo: cegando a los encuestados sobre las hipótesis

- ▶ Véase el estudio de Scacco y Warren sobre la teoría del contacto social en Kaduna (Nigeria), en donde los participantes fueron reclutados en un programa de conocimientos informáticos, pero que no se anunció como programa para reducir los prejuicios y la discriminación.

¿Cuánto error es demasiado?

- ▶ Reducir el error de medición es importante, pero puede ser bastante costoso. Entonces, ¿cuánto hay que hacer?
- ▶ Depende de la escala y de sus objetivos.
 - ▶ Comparar el tamaño de los errores con el tamaño del efecto del tratamiento. Cambiar de actitud es difícil. El sesgo de deseabilidad social puede ser grande en comparación con los pequeños efectos del tratamiento sobre las actitudes.
 - ▶ Compare el tamaño de los errores con el rango posible de esa variable de resultado. Tener un céntimo menos en su saldo bancario no afecta de forma apreciable a nuestra medida sobre su riqueza total.

Consejos generales para la medición I

- ▶ Comience con la práctica estándar para sus indicadores, es decir, indicadores que la comunidad de investigadores haya acordado que representan el concepto de interés.
 - ▶ Estos indicadores habrán sido probados para usted, y la comparabilidad de las mediciones entre los estudios de investigación y los sitios es una virtud.
- ▶ Pero hay que ser cuidadoso y tener en cuenta si los indicadores estándar tienen sentido.
 - ▶ Por ejemplo, la forma de medir los ingresos puede variar entre zonas ricas y pobres.
- ▶ Del mismo modo, la forma de medir las actitudes políticas puede ser diferente en regímenes más o menos democráticos.

Consejos generales para la medición II

- ▶ Recordemos que las mediciones de habilidad matemática y verbal que estaban relacionadas con la navegación eran práctica habitual en los años 20 y 80 en EE.UU. Recuerde que está midiendo constructos sociales y debe prestar atención, en la medida de lo posible, a las creencias previas, parcialmente o no examinadas, que usted y la comunidad de investigadores aportan al estudio (para algunos ejemplos más vívidos de esta cuestión véase Gould ([1981] 1996)).
- ▶ Conecte con otros investigadores de su área temática.
- ▶ Céntrese en una medición más delicada en el rango de la variable en el que espera el cambio.

Recursos para la investigación de encuestas y artículos de encuestas

- ▶ Los principales archivos de datos se pueden buscar por temas. Un archivo importante es el ICPSR:

<https://www.icpsr.umich.edu/web/pages/ICPSR/>

- ▶ El Pew Research Center Sobre el Diseño de Cuestionarios

<https://www.pewresearch.org/methods/u-s-survey-research/questionnaire-design/>

- ▶ Tools4Dev Para Escribir Preguntas

<http://www.tools4dev.org/resources/how-to-write-awesome-survey-questions-part-1/>

- ▶ Las bibliotecas universitarias suelen tener guías de recursos de datos:

<https://guides.libraries.emory.edu/c.php?g=944707&p=6810109>

Referencias

[10 cosas que debe saber sobre la medición en experimentos(<https://egap.org/resource/10-cosas-que-debe-saber-sobre-la-medicion-en-experimentos/>)

Adcock, Robert, and David Collier. 2001. "Measurement Validity: A Shared Standard for Qualitative and Measurement Validity: A Shared Standard for Qualitative and Quantitative Research." *American Political Science Review* 95 (3): 529–46.

Gould, Stephen Jay. [1981] 1996. *The Mismeasure of Man*. WW Norton & Company.

Shadish, William R., Thomas D. Cook, Donald Thomas Campbell, et al. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*/William r. Shadish, Thomas d. Cook, Donald t. Campbell. Boston: Houghton Mifflin,.