

# 02



TALLER DE INVESTIGACIÓN APLICADA

## Diseño de instrumentos y calidad en la medición

Pablo Argote

Clase 2 · Mayo 2026 · Miércoles 13 · 6–9 PM

# Recap: Clase 1



*Del brief al diseño investigable*

- 01 Leer el brief** Lo que el cliente dice y lo que falta. La inquietud rara vez llega en formato investigable.
- 02 Traducir a propuesta** Aclarar la decisión, distinguir lo importante de lo accesorio, detectar qué información falta.
- 03 Caso Banco X** Relanzamiento de producto financiero → 3 objetivos: segmentación, atributos, validación de claims.
- 04 Lo que NO hacer** Partir por la metodología, sobreofrecer, asumir que más detalle es más sofisticación.

*Hoy bajamos al diseño: **a quién** preguntar, **qué** preguntar, y **cómo** preguntar mejor.*



HOY

# Pregunta de la clase

*¿Cómo traducir un concepto en preguntas que efectivamente capturen la decisión que importa?*

## BLOQUE 1

### Muestreo

Probabilístico vs. targeteado · pesos del censo · cuándo sí, cuándo no

## BLOQUE 2

### Preguntas

Operacionalización · escalas · relevancia vs. posición · validación de claims

## BLOQUE 3

### Experimentos

Vignette · list experiment · conjoint · MaxDiff

*\* Habrá 2 breaks de 10 min*

# 01



BLOQUE 1

## Muestreo: ¿a quién le preguntamos?

# El ideal: muestreo probabilístico



Muestreo · 1 de 9

## DEFINICIÓN

Cada unidad de la población tiene una probabilidad **conocida y mayor a cero** de ser seleccionada. Es la base de la inferencia estadística clásica.

- **Aleatoriedad simple:** cada unidad con la misma probabilidad; requiere marco muestral completo.

## EJEMPLOS DEL GOLD STANDARD

Chile: **CASEN** (socioeconómica, MDS) y **CEP** (opinión pública). Colombia: **ECV del DANE**. Requieren marco muestral, presupuesto y tiempo — rara vez viables fuera del Estado o la academia.

# El ideal: muestreo probabilístico



Muestreo · 1 de 9

## DEFINICIÓN

Cada unidad de la población tiene una probabilidad **conocida y mayor a cero** de ser seleccionada. Es la base de la inferencia estadística clásica.

- **Aleatoria simple:** cada unidad con la misma probabilidad; requiere marco muestral completo.
- **Estratificada:** se divide en grupos (estratos) y se muestrea dentro de cada uno.

## EJEMPLOS DEL GOLD STANDARD

Chile: **CASEN** (socioeconómica, MDS) y **CEP** (opinión pública). Colombia: **ECV del DANE**. Requieren marco muestral, presupuesto y tiempo — rara vez viables fuera del Estado o la academia.

# El ideal: muestreo probabilístico



Muestreo · 1 de 9

## DEFINICIÓN

Cada unidad de la población tiene una probabilidad **conocida y mayor a cero** de ser seleccionada. Es la base de la inferencia estadística clásica.

- **Aleatoria simple:** cada unidad con la misma probabilidad; requiere marco muestral completo.
- **Estratificada:** se divide en grupos (estratos) y se muestrea dentro de cada uno.
- **Por conglomerados:** se eligen “racimos” (manzanas, escuelas, hogares) y se entrevista a sus miembros.

## EJEMPLOS DEL GOLD STANDARD

Chile: **CASEN** (socioeconómica, MDS) y **CEP** (opinión pública). Colombia: **ECV del DANE**. Requieren marco muestral, presupuesto y tiempo — rara vez viables fuera del Estado o la academia.

# La realidad: muestras targeteadas



Muestreo · 2 de 9

## OBSERVACIÓN

En investigación de mercado, el muestreo probabilístico casi nunca ocurre. La excepción son los estudios exploratorios que sí requieren una muestra representativa.

## LO HABITUAL — muestras intencionadas

- **Paneles online:** usuarios que se inscribieron por incentivo; sesgan a quien tiene tiempo y conexión.

## PREGUNTA SIEMPRE: ¿CÓMO SE GENERAN LOS DATOS?

¿Hay algo de aleatoriedad en el contacto? ¿O siempre responden los mismos? Si el proveedor recicla a los mismos usuarios, los sesgos quedan fijos en cada estudio.

# La realidad: muestras targeteadas



Muestreo · 2 de 9

## OBSERVACIÓN

En investigación de mercado, el muestreo probabilístico casi nunca ocurre. La excepción son los estudios exploratorios que sí requieren una muestra representativa.

## LO HABITUAL — muestras intencionadas

- **Paneles online:** usuarios que se inscribieron por incentivo; sesgan a quien tiene tiempo y conexión.
- **Base de clientes:** CRM propio; útil pero solo cubre a quienes ya son clientes.

## PREGUNTA SIEMPRE: ¿CÓMO SE GENERAN LOS DATOS?

¿Hay algo de aleatoriedad en el contacto? ¿O siempre responden los mismos? Si el proveedor recicla a los mismos usuarios, los sesgos quedan fijos en cada estudio.

# La realidad: muestras targeteadas



Muestreo · 2 de 9

## OBSERVACIÓN

En investigación de mercado, el muestreo probabilístico casi nunca ocurre. La excepción son los estudios exploratorios que sí requieren una muestra representativa.

## LO HABITUAL — muestras intencionadas

- **Paneles online:** usuarios que se inscribieron por incentivo; sesgan a quien tiene tiempo y conexión.
- **Base de clientes:** CRM propio; útil pero solo cubre a quienes ya son clientes.
- **Snowball y referidos:** útil en targets difíciles, pero replican el mundo del primer contacto.

## PREGUNTA SIEMPRE: ¿CÓMO SE GENERAN LOS DATOS?

¿Hay algo de aleatoriedad en el contacto? ¿O siempre responden los mismos? Si el proveedor recicla a los mismos usuarios, los sesgos quedan fijos en cada estudio.

# ¿Qué es ponderar?



Muestreo · 3 de 9

## DEFINICIÓN

Ajuste **post-hoc** (después de recolectar) que asigna a cada usuario un peso  $w_g$  para que la muestra refleje la distribución conocida de la población.

## REQUISITO

Necesita un **referente externo** con la distribución real: censo, registros administrativos. Sin eso, no hay contra qué ajustar.

## FÓRMULA

**Peso** (una variable):

$$w_g = \frac{P_g^{\text{pob}}}{\hat{p}_g^{\text{mue}}}$$

proporción **poblacional** sobre proporción **muestral** en el grupo  $g$

**Rake** (varias variables):

$$w_i = \prod_v \frac{P_v}{\hat{p}_v}$$

**producto** de los pesos por cada variable  $v$ , iterando hasta calzar todas las marginales

# ¿Ponderar al censo? — caso Banco X



Muestreo · 4 de 9

## SITUACIÓN

Estudio sobre clientes del Banco X. Ya son ~30 % de la población. Un cliente pregunta:  
“¿Aplicamos pesos del censo (edad, sexo, NSE) para que los resultados sean representativos?”

## POR QUÉ NO ES OBVIO

El target ya está concentrado demográficamente: clientes Banco X tienden a ser más urbanos, más jóvenes y de mayor SES. Esas son justamente las variables con las que el cliente quiere ponderar. Ajustar por algo que ya está **implicado** en la definición del target agrega poco.

## LA REGLA QUE EMERGE

Mientras más correlacionada está una variable demográfica con la que define al target, **menos necesario** es ponderar por esa variable demográfica.

# Cuándo sí ponderar



Muestreo · 5 de 9

## CASO — Banco X en estudio amplio

Estudio cubre **toda la población** de consumidores. Banco X queda sub-representado: **10 % en muestra, 30 % en realidad.**

### DOS OPCIONES

#### A) PONDERAR POR DEMOGRÁFICOS

Edad, urbanidad, SES (variables correlacionadas con ser cliente de Banco X). Rake weights con varias marginales del censo.

*Asume que esas variables explican el sub-muestreo.*

#### B) PONDERAR DIRECTO POR SER CLIENTE BANCO X

Indicador binario “cliente Banco X”.  $w = P_g/p_g$ :

**Cientes Banco X:**  $0,30/0,10 = 3,0\times$

**No clientes:**  $0,70/0,90 \approx 0,78\times$

**CUIDADO:** pesos altos sobre pocas observaciones inflan los errores estándar. Ponderar mucho con poca data hace *ruido, no precisión.*

# Cuándo **SÍ** ayuda ponderar



Muestreo · 6 de 9

## CONDICIONES

- **Independencia:** La variable de ponderación es **independiente** de lo que define al target.

*La pregunta antes de ponderar: ¿esta variable agrega información que el target no me da ya?*

# Cuándo **SÍ** ayuda ponderar



Muestreo · 6 de 9

## CONDICIONES

- **Independencia:** La variable de ponderación es **independiente** de lo que define al target.
- **Género:** Casi siempre seguro: rara vez correlaciona con la conducta o segmento que define el target.

*La pregunta antes de ponderar: ¿esta variable agrega información que el target no me da ya?*

# Cuándo **SÍ** ayuda ponderar



Muestreo · 6 de 9

## CONDICIONES

- **Independencia:** La variable de ponderación es **independiente** de lo que define al target.
- **Género:** Casi siempre seguro: rara vez correlaciona con la conducta o segmento que define el target.
- **Universo amplio:** Hay referente externo confiable (censo, registros) y la población definida es general.

*La pregunta antes de ponderar: ¿esta variable agrega información que el target no me da ya?*

# Cuándo **SÍ** ayuda ponderar



Muestreo · 6 de 9

## CONDICIONES

- **Independencia:** La variable de ponderación es **independiente** de lo que define al target.
- **Género:** Casi siempre seguro: rara vez correlaciona con la conducta o segmento que define el target.
- **Universo amplio:** Hay referente externo confiable (censo, registros) y la población definida es general.
- **Suficientes observaciones:** En cada celda hay  $n$  suficiente para que el peso no infle el error estándar.

*La pregunta antes de ponderar: ¿esta variable agrega información que el target no me da ya?*

# Cuándo **NO** ayuda — ejemplo



Muestreo · 7 de 9

## SETUP

**Target:** clientes Banco X. Decidimos ponderar por **educación**.

## LOS DATOS — “educación media como máximo”

Censo (Chile): **40 %**

Muestra: **10 %**

Peso aplicado: **4×**

## EL ERROR

Ese **40 %** describe a *toda* la población — no a los clientes Banco X. Dentro del target, la proporción con educación media **no sabemos cuál es**, y probablemente no es 40 %. Estamos ajustando con un dato que no aplica.

# ¿Cuándo sirven de verdad los pesos?



Muestreo · 8 de 9

## LA IDEA

Ponderar por edad, sexo y NSE arregla desvíos de **mi muestra real vs. la muestra que pensé**.

También corrige **no-respuesta** cuando la demografía correlaciona con el outcome (ej.: jóvenes responden menos y opinan distinto).

Ej: encuesta de aprobación presidencial — ahí ponderar es **clave**, quiero ver al país completo.

## NO SIRVE CUANDO

- El target es **específico** (no es “el país”).

# ¿Cuándo sirven de verdad los pesos?



Muestreo · 8 de 9

## LA IDEA

Ponderar por edad, sexo y NSE arregla desvíos de **mi muestra real vs. la muestra que pensé**.

También corrige **no-respuesta** cuando la demografía correlaciona con el outcome (ej.: jóvenes responden menos y opinan distinto).

Ej: encuesta de aprobación presidencial — ahí ponderar es **clave**, quiero ver al país completo.

## NO SIRVE CUANDO

- El target es **específico** (no es “el país”).
- El target está **correlacionado** con la variable demográfica.

# ¿Cuándo sirven de verdad los pesos?



Muestreo · 8 de 9

## LA IDEA

Ponderar por edad, sexo y NSE arregla desvíos de **mi muestra real vs. la muestra que pensé**.

También corrige **no-respuesta** cuando la demografía correlaciona con el outcome (ej.: jóvenes responden menos y opinan distinto).

Ej: encuesta de aprobación presidencial — ahí ponderar es **clave**, quiero ver al país completo.

## NO SIRVE CUANDO

- El target es **específico** (no es “el país”).
- El target está **correlacionado** con la variable demográfica.
- **No sabemos** cómo se distribuyen las demografías **dentro** del target.

# Preguntas antes de ponderar



Muestreo · 9 de 9

## a) ¿Cuánto me desvié de mi idea en la práctica?

Si hay desvío, sí. *Ej:* quería 30 % Banco X, llegué a 27 % → pondero por presencia de Banco X, nada más.

# Preguntas antes de ponderar



Muestreo · 9 de 9

## a) ¿Cuánto me desvié de mi idea en la práctica?

Si hay desvío, sí. *Ej:* quería 30 % Banco X, llegué a 27 % → pondero por presencia de Banco X, nada más.

## b) ¿Es el target muy específico?

Mejor no.

# Preguntas antes de ponderar



Muestreo · 9 de 9

## a) ¿Cuánto me desvié de mi idea en la práctica?

Si hay desvío, sí. *Ej:* quería 30 % Banco X, llegué a 27 % → pondero por presencia de Banco X, nada más.

## b) ¿Es el target muy específico?

Mejor no.

## c) ¿Están las variables demográficas correlacionadas con el target?

Mejor no.

# Preguntas antes de ponderar



Muestreo · 9 de 9

## a) ¿Cuánto me desvié de mi idea en la práctica?

Si hay desvío, sí. *Ej: quería 30 % Banco X, llegué a 27 % → pondero por presencia de Banco X, nada más.*

## b) ¿Es el target muy específico?

Mejor no.

## c) ¿Están las variables demográficas correlacionadas con el target?

Mejor no.

## d) ¿Conozco la distribución de las demográficas dentro del target?

Si no las conozco, o no tengo ninguna información, mejor no.

# 02



BLOQUE 2

## Preguntas: de conceptos a ítems

*Una pincelada — el diseño de instrumentos es un campo entero. La idea fuerza: **las mejores preguntas obligan a priorizar**, no a evaluar en abstracto.*

# ¿Qué es operacionalizar?



Operacionalización · 1 de 4

## DEFINICIÓN

Operacionalizar es **traducir un concepto abstracto en variables observables y medibles.**

- Los conceptos que importan en investigación rara vez se miden **directamente.**

*Sin operacionalización, el instrumento mide cualquier cosa menos lo que queremos saber.*

# ¿Qué es operacionalizar?



Operacionalización · 1 de 4

## DEFINICIÓN

Operacionalizar es **traducir un concepto abstracto en variables observables y medibles**.

- Los conceptos que importan en investigación rara vez se miden **directamente**.
- “Satisfacción”, “confianza”, “lealtad”, “intención de compra” no vienen con una **unidad de medida obvia**.

*Sin operacionalización, el instrumento mide cualquier cosa menos lo que queremos saber.*

# ¿Qué es operacionalizar?



Operacionalización · 1 de 4

## DEFINICIÓN

Operacionalizar es **traducir un concepto abstracto en variables observables y medibles**.

- Los conceptos que importan en investigación rara vez se miden **directamente**.
- “Satisfacción”, “confianza”, “lealtad”, “intención de compra” no vienen con una **unidad de medida obvia**.
- Operacionalizar es el **punto** entre la pregunta de investigación y los datos.

*Sin operacionalización, el instrumento mide cualquier cosa menos lo que queremos saber.*

# No es trivial



Operacionalización · 2 de 4

## PREGUNTA DE INVESTIGACIÓN

*¿Qué tan satisfechos están los usuarios con el servicio?*

- “En una escala de 1 a 10, ¿cuán satisfecho está con el servicio?”

*escala global*

## OBSERVACIÓN

Cada operacionalización captura algo distinto. Elegir mal el indicador puede llevar a conclusiones equivocadas con datos perfectamente recolectados.

# No es trivial



Operacionalización · 2 de 4

## PREGUNTA DE INVESTIGACIÓN

¿Qué tan satisfechos están los usuarios con el servicio?

- “En una escala de 1 a 10, ¿cuán satisfecho está con el servicio?” *escala global*
- “¿Qué tan probable es que recomiende el servicio? (0-10)” *NPS*

## OBSERVACIÓN

Cada operacionalización captura algo distinto. Elegir mal el indicador puede llevar a conclusiones equivocadas con datos perfectamente recolectados.

# No es trivial



Operacionalización · 2 de 4

## PREGUNTA DE INVESTIGACIÓN

¿Qué tan satisfechos están los usuarios con el servicio?

- “En una escala de 1 a 10, ¿cuán satisfecho está con el servicio?” *escala global*
- “¿Qué tan probable es que recomiende el servicio? (0–10)” *NPS*
- “¿Volvería a contratar el servicio?” *intención de comportamiento*

## OBSERVACIÓN

Cada operacionalización captura algo distinto. Elegir mal el indicador puede llevar a conclusiones equivocadas con datos perfectamente recolectados.

# No es trivial



Operacionalización · 2 de 4

## PREGUNTA DE INVESTIGACIÓN

¿Qué tan satisfechos están los usuarios con el servicio?

- “En una escala de 1 a 10, ¿cuán satisfecho está con el servicio?” *escala global*
- “¿Qué tan probable es que recomiende el servicio? (0–10)” *NPS*
- “¿Volvería a contratar el servicio?” *intención de comportamiento*
- “¿Cuántas veces ha contactado al servicio en los últimos 6 meses?” *conducta revelada*

## OBSERVACIÓN

Cada operacionalización captura algo distinto. Elegir mal el indicador puede llevar a conclusiones equivocadas con datos perfectamente recolectados.

# No es trivial



Operacionalización · 2 de 4

## PREGUNTA DE INVESTIGACIÓN

¿Qué tan satisfechos están los usuarios con el servicio?

- “En una escala de 1 a 10, ¿cuán satisfecho está con el servicio?” *escala global*
- “¿Qué tan probable es que recomiende el servicio? (0–10)” *NPS*
- “¿Volvería a contratar el servicio?” *intención de comportamiento*
- “¿Cuántas veces ha contactado al servicio en los últimos 6 meses?” *conducta revelada*
- “¿Cómo describiría su última experiencia?” *narrativa abierta*

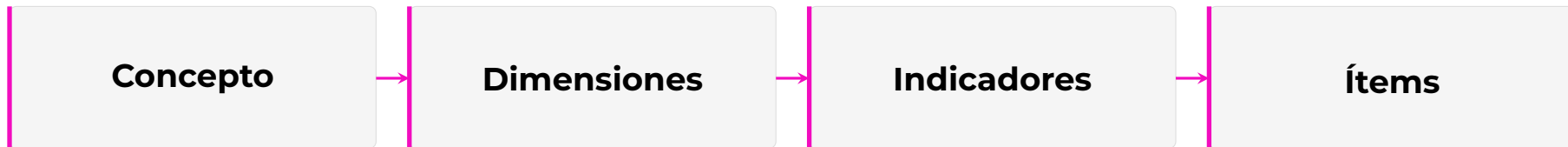
## OBSERVACIÓN

Cada operacionalización captura algo distinto. Elegir mal el indicador puede llevar a conclusiones equivocadas con datos perfectamente recolectados.

# El pipeline



Operacionalización · 3 de 4



\* **Dimensiones** es opcional: si el concepto ya es concreto (ej. “frecuencia de compra”), se salta directo a indicadores.

## EJEMPLO — “Confianza en la marca”

**Dimensiones**      fiabilidad · honestidad · cercanía

**Indicadores**      percepción de cumplimiento · transparencia · trato

**Ítems**      “¿Cuán de acuerdo está con que la marca cumple lo que promete?” (Likert 1–5)

# Tipos de preguntas



Operacionalización · 4 de 4

## CERRADAS — para medir

- 01 Sí o No (dicotómica)
- 02 Opción múltiple (única o múltiple)
- 03 Likert (acuerdo o frecuencia)
- 04 Diferencial semántico (dos polos)
- 05 Ranking (priorización)
- 06 Numérica (edad, ingreso, NPS)

## ABIERTAS — para explorar

- 01 Respuesta corta libre
- 02 Respuesta extendida
- 03 Asociación espontánea
- 04 Completación de frases

*Las abiertas dan riqueza, pero son **engorrosas**: caras de codificar, difíciles de comparar y muy ruidosas. Úsalas con cuidado.*

# Ejemplo: Sí o No



*Cuándo funciona y cuándo no*

## **MALA PREGUNTA**

*“¿Cree usted que se debe cuidar el medio ambiente?”*

→ 95 % dice “sí”. Mide **deseabilidad social pura**, no aporta información para decidir.

## **BUENA PREGUNTA**

*“¿Recicló usted en su hogar al menos una vez la semana pasada?”*

→ Conducta concreta y verificable. Discrimina entre quienes lo hacen y quienes no.

*Sí/No solo sirve para **hechos verificables**, no para opiniones generales.*

# Ejemplo: Likert



Actitudes graduales

## LA PREGUNTA

“¿Qué tan de acuerdo está con la frase: esta marca cumple lo que promete?”

## ESCALA



1

muy en desacuerdo



2



3



4



5

muy de acuerdo

## CUÁNDO USAR — Y EL LÍMITE

Útil para actitudes y opiniones graduales. **Pero comparte sesgos con la dicotómica:** si todos los ítems son “positivos”, la gente responde 4 o 5 a casi todo, sin discriminar.

# Diferencial semántico



*Dos polos con punto medio*

## CÓMO SE VE

Se presentan dos adjetivos opuestos y el usuario se ubica entre ambos.

<b>Confiable</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<b>Poco confiable</b>
<b>Moderno</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<b>Anticuado</b>
<b>Caro</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<b>Económico</b>

## EJEMPLO — AFP

“¿Aprueba las AFP?” → “no, son los malos”. Con diferencial *100 % estatal* ↔ *100 % privado*: la gente se ubica al medio — aparece el matiz real.

- Captura **matiz/gradiente** entre dos polos — ideal para percepciones de marca y posicionamiento.

# Diferencial semántico



*Dos polos con punto medio*

## CÓMO SE VE

Se presentan dos adjetivos opuestos y el usuario se ubica entre ambos.

<b>Confiable</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<b>Poco confiable</b>
<b>Moderno</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<b>Anticuado</b>
<b>Caro</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<b>Económico</b>

## EJEMPLO — AFP

“¿Aprueba las AFP?” → “no, son los malos”. Con diferencial *100 % estatal* ↔ *100 % privado*: la gente se ubica al medio — aparece el matiz real.

- Captura **matiz/gradiente** entre dos polos — ideal para percepciones de marca y posicionamiento.
- El punto medio no es “no sé”: es indiferencia o equidistancia.

# Diferencial semántico



*Dos polos con punto medio*

## CÓMO SE VE

Se presentan dos adjetivos opuestos y el usuario se ubica entre ambos.

<b>Confiable</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<b>Poco confiable</b>
<b>Moderno</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<b>Anticuado</b>
<b>Caro</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<b>Económico</b>

## EJEMPLO — AFP

“¿Aprueba las AFP?” → “no, son los malos”. Con diferencial *100 % estatal* ↔ *100 % privado*: la gente se ubica al medio — aparece el matiz real.

- Captura **matiz/gradiente** entre dos polos — ideal para percepciones de marca y posicionamiento.
- El punto medio no es “no sé”: es indiferencia o equidistancia.
- Si los adjetivos no son verdaderos opuestos, la escala se rompe.

# Ejemplo: Ranking



Forzar a priorizar — la joya

## LA PREGUNTA

“Ordene de **más** a **menos** importante al elegir un banco:”

1. Tasa preferencial
2. Atención rápida en sucursal
3. Plataforma digital fácil de usar
4. Sucursales cercanas

El usuario no puede decir “todos son importantes”. Tiene que **ceder algo** — y eso es lo que revela qué pesa de verdad.

# Posición vs. relevancia



*Una distinción clave*

## POSICIÓN

- ¿Qué **piensa** el usuario?
- ¿Está a favor o en contra?
- ¿Qué tan de acuerdo está?
- Se mide bien con **Likert**.

## RELEVANCIA

- ¿Cuánto le **importa de verdad**?
- ¿Pesa en su **decisión**?
- ¿**Sacrificaría** otra cosa por esto?
- Se mide bien con **priorización forzada**.

### LA TRAMPA — ejemplo, ciencia política

“¿Le importa el orden público?” → casi todos dicen que sí.

“¿Orden público o libertad individual?” → **revela qué pesa de verdad**.

Medir solo posición confunde “importante en abstracto” con “decisivo en la práctica”.

# Ejemplo intuitivo



Posición vs. relevancia · caso ilustrativo

## PREGUNTA 1 — POSICIÓN

“¿Está de acuerdo con que la sustentabilidad ambiental de una empresa es importante?”

*Respuesta típica: ~90 % “muy de acuerdo”.*

## PREGUNTA 2 — RELEVANCIA

“Cuando elige entre dos productos similares, ¿qué pesa más en su decisión: precio, marca, sustentabilidad o conveniencia?”

*Respuesta típica: sustentabilidad queda 4ª, con ~5 %.*

La gente **piensa** que la sustentabilidad importa (posición). Pero al **decidir**, compra por precio (relevancia baja).

# Preguntas mal formuladas



Anti-ejemplos · 1 de 2

“¿Cree usted que se debe cuidar el medio ambiente?”

*respuesta obvia (~95, % sí); no aporta información, mide deseabilidad social pura.*

“¿Está usted de acuerdo con la gestión del gobierno?”

*¿qué gestión? ¿en qué área? mide aprobación general, no la decisión que importa.*

“¿Cree que el país va por buen camino?”

*“buen camino” significa cosas distintas para distintas personas.*

“¿No le parece preocupante el alza de la delincuencia?”

*pregunta cargada con doble negación; sesga hacia “sí”.*

“¿Aprobaría una reforma tributaria que aumente la recaudación para programas sociales?”

*premisa cargada; cambia si se describe como “aumento de impuestos”.*

# Preguntas mal formuladas



*Anti-ejemplos · 2 de 2*

“¿Qué importancia le asigna a la educación, la salud y la seguridad?”

*mide posición (todo importa); no fuerza priorización, no sirve para decidir.*

“¿Cuán satisfecho está con su vida en general?”

*tan global que no se puede accionar; mejor preguntar por dimensiones específicas.*

“¿Recomendaría este producto a sus amigos y familiares?”

*doble pregunta (¿amigos? ¿familiares?). NPS resuelve preguntando solo “a un amigo”.*

“En su opinión, ¿los inmigrantes contribuyen positivamente al país?”

*tema sensible → deseabilidad social fuerte; mejor usar viñeta o list experiment.*

*Patrón común: miden posición sin captar relevancia, cargan la respuesta, o son demasiado vagas para accionar.*

# Caso Banco X — aplicación



*Validar claims para un producto financiero*

## CONTEXTO — 4 claims a validar

- “Tasa preferencial garantizada por 12 meses.”
- “Sin costos de mantención el primer año.”
- “Aprobación 100 % online en menos de 24 horas.”
- “Respaldado por el banco más grande del país.”

## PROBLEMA — Likert

“¿Cuán importante es X?” → los 4 claims salen **entre 4 y 5**. Imposible priorizar.

## SOLUCIÓN — forzar a priorizar

Pedir al usuario que elija **cuál claim importa más** y cuál menos. La elección forzada revela la prioridad real.

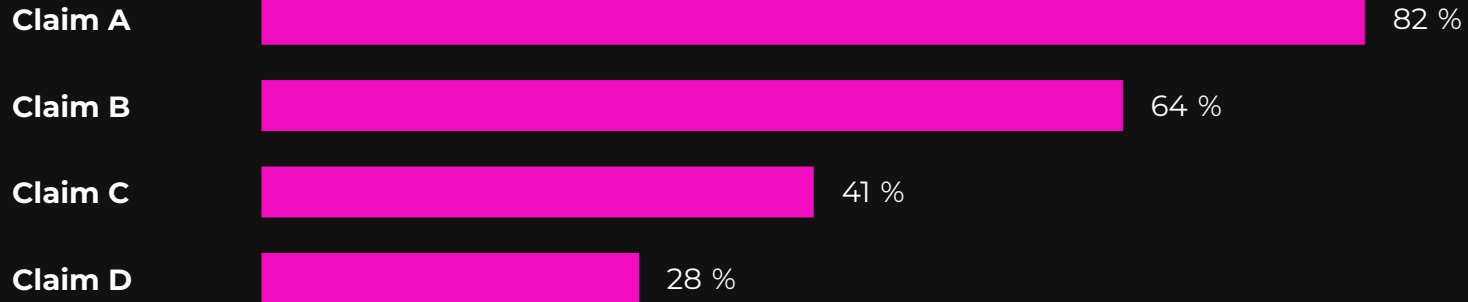
*Mismo patrón que las AFP: el sí/no o Likert directo aplasta el matiz; el trade-off captura qué importa de verdad.*

# El resultado en Banco X



*Validación de claims · priorización forzada*

## IMPORTANCIA RELATIVA



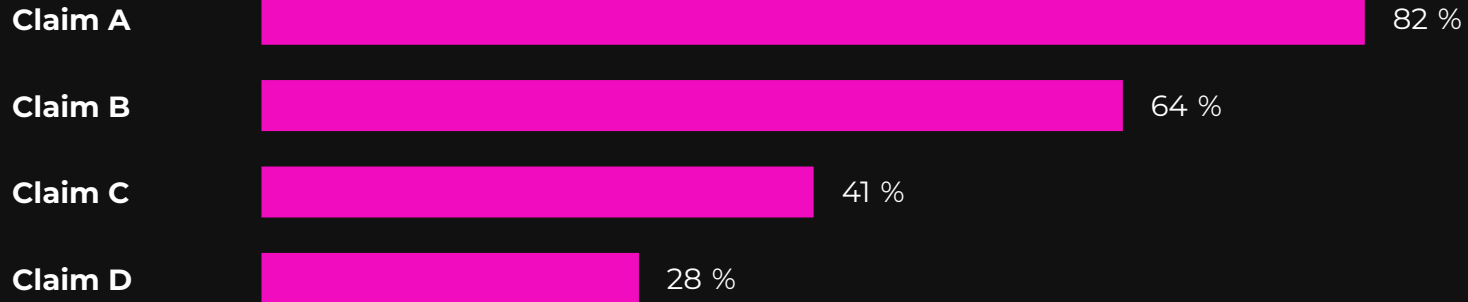
- Si hubiéramos preguntado solo en Likert, los cuatro habrían quedado entre 4 y 5.

# El resultado en Banco X



Validación de claims · priorización forzada

## IMPORTANCIA RELATIVA



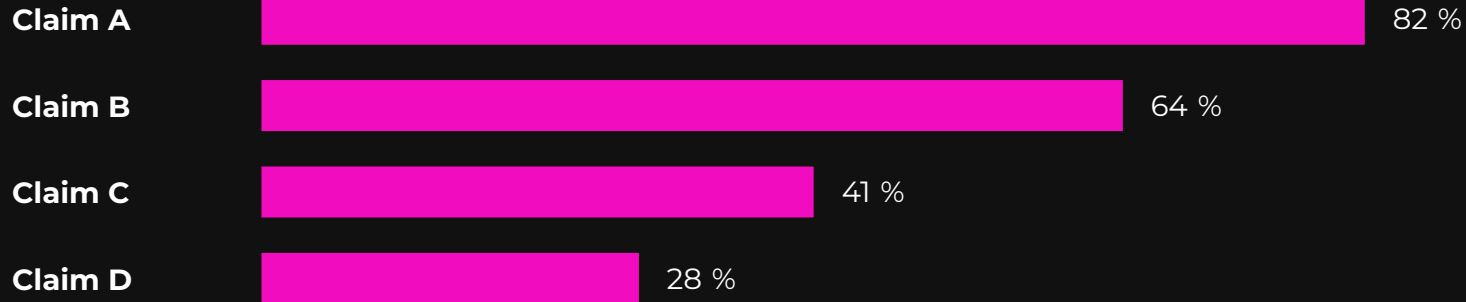
- Si hubiéramos preguntado solo en Likert, los cuatro habrían quedado entre 4 y 5.
- Forzar la priorización nos permitió ordenar y descartar con criterio cuantitativo.

# El resultado en Banco X



Validación de claims · priorización forzada

## IMPORTANCIA RELATIVA



- Si hubiéramos preguntado solo en Likert, los cuatro habrían quedado entre 4 y 5.
- Forzar la priorización nos permitió ordenar y descartar con criterio cuantitativo.
- La decisión de comunicación se basó en el Claim A, no en intuición.

# 03



## BLOQUE 3

# Experimentos para mejorar la medición

*Para reducir sesgos y captar lo que importa — **no** para estudiar relaciones causa-efecto. La causalidad real no se resuelve dentro de una encuesta.*

# ¿Por qué no basta con preguntar?



*Sesgos de las preguntas directas*

## EL PROBLEMA

Aun con buena operacionalización y escalas correctas, las preguntas directas tienen **sesgos sistemáticos**.

- **Deseabilidad social:** se reporta lo socialmente aceptable, no lo real.

*La salida es captar preferencias de forma **indirecta**, vía experimentos dentro de la encuesta.*

# ¿Por qué no basta con preguntar?



*Sesgos de las preguntas directas*

## EL PROBLEMA

Aun con buena operacionalización y escalas correctas, las preguntas directas tienen **sesgos sistemáticos**.

- **Deseabilidad social:** se reporta lo socialmente aceptable, no lo real.
- **Anclaje y framing:** la respuesta cambia con cómo se presenta la pregunta.

*La salida es captar preferencias de forma **indirecta**, vía experimentos dentro de la encuesta.*

# ¿Por qué no basta con preguntar?



*Sesgos de las preguntas directas*

## EL PROBLEMA

Aun con buena operacionalización y escalas correctas, las preguntas directas tienen **sesgos sistemáticos**.

- **Deseabilidad social:** se reporta lo socialmente aceptable, no lo real.
- **Anclaje y framing:** la respuesta cambia con cómo se presenta la pregunta.
- **Falta de introspección:** el usuario no siempre sabe por qué decide lo que decide.

*La salida es captar preferencias de forma **indirecta**, vía experimentos dentro de la encuesta.*

# ¿Por qué no basta con preguntar?



*Sesgos de las preguntas directas*

## EL PROBLEMA

Aun con buena operacionalización y escalas correctas, las preguntas directas tienen **sesgos sistemáticos**.

- **Deseabilidad social:** se reporta lo socialmente aceptable, no lo real.
- **Anclaje y framing:** la respuesta cambia con cómo se presenta la pregunta.
- **Falta de introspección:** el usuario no siempre sabe por qué decide lo que decide.
- **Trade-offs invisibles:** “todo me importa” cuando se pregunta una cosa a la vez.

*La salida es captar preferencias de forma **indirecta**, vía experimentos dentro de la encuesta.*

# Captar preferencias indirectamente



4 técnicas que veremos hoy

## IDEA CENTRAL

En lugar de preguntar “¿qué le importa?”, se construye una situación donde el usuario **revela** sus preferencias al **elegir, evaluar o decidir**. **Objetivo: mejor medición, no inferencia causal.**

### 01 Vignette

*escenarios randomizados*

### 02 List experiment

*prevalencia de conductas sensibles*

### 03 Conjoint

*trade-offs entre múltiples atributos*

### 04 MaxDiff

*priorización forzada entre alternativas*



## QUÉ ES

Se presenta un escenario corto cuyos atributos **varían aleatoriamente** entre usuarios. La diferencia en las respuestas estima el **efecto del atributo** en la evaluación.

## VERSIÓN A

“**María**, ingeniera, 32 años, egresada de universidad **pública**, 5 años de experiencia. ¿Qué tan apta para el cargo?”

## VERSIÓN B

“**Pedro**, ingeniero, 32 años, egresado de universidad **privada**, 5 años de experiencia. ¿Qué tan apto para el cargo?”

- Se aísla el efecto de género y tipo de universidad.

## QUÉ ES

Se presenta un escenario corto cuyos atributos **varían aleatoriamente** entre usuarios. La diferencia en las respuestas estima el **efecto del atributo** en la evaluación.

## VERSIÓN A

“**María**, ingeniera, 32 años, egresada de universidad **pública**, 5 años de experiencia. ¿Qué tan apta para el cargo?”

## VERSIÓN B

“**Pedro**, ingeniero, 32 años, egresado de universidad **privada**, 5 años de experiencia. ¿Qué tan apto para el cargo?”

- Se aísla el efecto de género y tipo de universidad.
- El usuario no sabe que existe otra versión → reduce deseabilidad social.

## QUÉ ES

Se presenta un escenario corto cuyos atributos **varían aleatoriamente** entre usuarios. La diferencia en las respuestas estima el **efecto del atributo** en la evaluación.

## VERSIÓN A

“**María**, ingeniera, 32 años, egresada de universidad **pública**, 5 años de experiencia. ¿Qué tan apta para el cargo?”

## VERSIÓN B

“**Pedro**, ingeniero, 32 años, egresado de universidad **privada**, 5 años de experiencia. ¿Qué tan apto para el cargo?”

- Se aísla el efecto de género y tipo de universidad.
- El usuario no sabe que existe otra versión → reduce deseabilidad social.
- Útil cuando: hay sospecha de sesgo, prejuicio, o el atributo es difícil de declarar.

# Deseabilidad social



*Cuando la respuesta no coincide con la conducta*

## EL SESGO

El usuario reporta lo **socialmente aceptable**, no lo que efectivamente piensa o hace.

## CASO CLÁSICO — participación electoral en Chile

CEP / encuestas pre-electorales: ~70 % declara que va a votar.

Servel · participación real: ~45 % efectivamente vota.

*La diferencia (~25 pp) no es error muestral: es deseabilidad social. Votar es “lo correcto”.*

- Otros ejemplos: reciclaje, donación, lectura, ejercicio, consumo de alcohol.

# Deseabilidad social



*Cuando la respuesta no coincide con la conducta*

## EL SESGO

El usuario reporta lo **socialmente aceptable**, no lo que efectivamente piensa o hace.

## CASO CLÁSICO — participación electoral en Chile

CEP / encuestas pre-electorales: ~70 % declara que va a votar.

Servel · participación real: ~45 % efectivamente vota.

*La diferencia (~25 pp) no es error muestral: es deseabilidad social. Votar es “lo correcto”.*

- Otros ejemplos: reciclaje, donación, lectura, ejercicio, consumo de alcohol.
- Mientras más normativa la conducta, mayor el sesgo.

# Deseabilidad social



*Cuando la respuesta no coincide con la conducta*

## EL SESGO

El usuario reporta lo **socialmente aceptable**, no lo que efectivamente piensa o hace.

## CASO CLÁSICO — participación electoral en Chile

CEP / encuestas pre-electorales: ~70 % declara que va a votar.

Servel · participación real: ~45 % efectivamente vota.

*La diferencia (~25 pp) no es error muestral: es deseabilidad social. Votar es “lo correcto”.*

- Otros ejemplos: reciclaje, donación, lectura, ejercicio, consumo de alcohol.
- Mientras más normativa la conducta, mayor el sesgo.
- No se elimina con buena redacción: hay que diseñar la medición de otra manera — como el **list experiment** (siguiente).

# List experiment



Técnicas indirectas · 2 de 4

## QUÉ ES

Técnica para medir la **prevalencia** de una conducta u opinión sensible **sin que el usuario revele individualmente** su respuesta.

## MECÁNICA — dos grupos randomizados

- **Control:** *“Indique cuántas de las siguientes le pasaron en el último año:”* + 3 actividades cotidianas (cambió de empleo, viajó al extranjero, donó a alguna causa).

## ESTIMACIÓN

Diferencia de medias entre grupos = prevalencia del ítem sensible. Útil cuando: discriminación, conductas ilegales, opiniones tabú.

# List experiment



Técnicas indirectas · 2 de 4

## QUÉ ES

Técnica para medir la **prevalencia** de una conducta u opinión sensible **sin que el usuario revele individualmente** su respuesta.

## MECÁNICA — dos grupos randomizados

- **Control:** *“Indique cuántas de las siguientes le pasaron en el último año:”* + 3 actividades cotidianas (cambió de empleo, viajó al extranjero, donó a alguna causa).
- **Tratamiento:** Las mismas 3 + el ítem sensible (ej.: **“recibió un pago de alguien vinculado al narcotráfico”**).

## ESTIMACIÓN

Diferencia de medias entre grupos = prevalencia del ítem sensible. Útil cuando: discriminación, conductas ilegales, opiniones tabú.

# List experiment



Técnicas indirectas · 2 de 4

## QUÉ ES

Técnica para medir la **prevalencia** de una conducta u opinión sensible **sin que el usuario revele individualmente** su respuesta.

## MECÁNICA — dos grupos randomizados

- **Control:** “Indique cuántas de las siguientes le pasaron en el último año:” + 3 actividades cotidianas (cambió de empleo, viajó al extranjero, donó a alguna causa).
- **Tratamiento:** Las mismas 3 + el ítem sensible (ej.: “**recibió un pago de alguien vinculado al narcotráfico**”).
- **Reporte:** El usuario solo dice **cuántas** le pasaron, no cuáles.

## ESTIMACIÓN

Diferencia de medias entre grupos = prevalencia del ítem sensible. Útil cuando: discriminación, conductas ilegales, opiniones tabú.

## QUÉ ES

Se presentan dos perfiles reales del producto y el usuario elige uno. La idea: la preferencia surge de una **elección concreta entre alternativas**, no de evaluar atributos en abstracto. Repitiendo, se estima la **preferencia por cada atributo por separado**.

## PRODUCTO FINANCIERO

Atributo	Opción 1	Opción 2
Tasa	4, %	7, %
Monto mínimo	\$500	\$5.000
Plazo	3 meses	12 meses
Aprobación	24 hrs	5 días

*¿Cuál preferiría contratar?*

*Cada elección es un trade-off real: prefiero tasa baja a costa de un plazo más largo. Eso revela el peso de cada atributo.*

## QUÉ ES

Se presentan subconjuntos de ítems (típicamente 4 a la vez); el usuario elige el más importante y el menos importante. Se repite varias veces.

## CLAIMS BANCO X

*De los siguientes beneficios, ¿cuál es el más importante y cuál el menos importante?*

- Tasa preferencial garantizada por 12 meses

*Estima utilidad relativa por ítem; escalable a 15–20 claims.*

## QUÉ ES

Se presentan subconjuntos de ítems (típicamente 4 a la vez); el usuario elige el más importante y el menos importante. Se repite varias veces.

## CLAIMS BANCO X

*De los siguientes beneficios, ¿cuál es el más importante y cuál el menos importante?*

- Tasa preferencial garantizada por 12 meses
- Sin costos de mantención el primer año

*Estima utilidad relativa por ítem; escalable a 15–20 claims.*

## QUÉ ES

Se presentan subconjuntos de ítems (típicamente 4 a la vez); el usuario elige el más importante y el menos importante. Se repite varias veces.

## CLAIMS BANCO X

*De los siguientes beneficios, ¿cuál es el más importante y cuál el menos importante?*

- Tasa preferencial garantizada por 12 meses
- Sin costos de mantención el primer año
- Aprobación 100,% online en menos de 24 horas

*Estima utilidad relativa por ítem; escalable a 15–20 claims.*

## QUÉ ES

Se presentan subconjuntos de ítems (típicamente 4 a la vez); el usuario elige el más importante y el menos importante. Se repite varias veces.

## CLAIMS BANCO X

*De los siguientes beneficios, ¿cuál es el más importante y cuál el menos importante?*

- Tasa preferencial garantizada por 12 meses
- Sin costos de mantención el primer año
- Aprobación 100,% online en menos de 24 horas
- Respaldo por el banco más grande del país

*Estima utilidad relativa por ítem; escalable a 15–20 claims.*

# ¿Cuándo usar cuál?



Decisión rápida · resumen

Técnica	Pregunta que responde	Cuándo
<b>Vignette</b>	¿Qué efecto tiene un atributo en una evaluación?	Sospecha de sesgo o prejuicio.
<b>List experiment</b>	¿Qué proporción hace o piensa X (sensible)?	Conductas tabú o no declarables.
<b>Conjoint</b>	¿Cuánto pesa cada atributo en la elección?	Diseño de productos con múltiples atributos.
<b>MaxDiff</b>	¿Qué importa más entre N alternativas?	Priorización de claims, mensajes, prioridades.

## PUNTO COMÚN

Todas combinan aleatorización (lógica experimental) con encuesta (representatividad).

# Cierre

---



*Lo que aprendimos · próximo paso*

## HOY

- Operacionalizar concepto → variable.

## PRÓXIMA CLASE

- Producción y análisis de datos.
- Software, IA, calidad de datos.
- Martes 19 de mayo · 6–9 PM

# Cierre

---



*Lo que aprendimos · próximo paso*

## HOY

- Operacionalizar concepto → variable.
- Tipos de preguntas: sí o no, Likert, ranking, diferencial semántico.

## PRÓXIMA CLASE

- Producción y análisis de datos.
- Software, IA, calidad de datos.
- Martes 19 de mayo · 6–9 PM

# Cierre

---



*Lo que aprendimos · próximo paso*

## HOY

- Operacionalizar concepto → variable.
- Tipos de preguntas: sí o no, Likert, ranking, diferencial semántico.
- Distinguir **posición** de **relevancia**.

## PRÓXIMA CLASE

- Producción y análisis de datos.
- Software, IA, calidad de datos.
- Martes 19 de mayo · 6–9 PM

# Cierre



*Lo que aprendimos · próximo paso*

## HOY

- Operacionalizar concepto → variable.
- Tipos de preguntas: sí o no, Likert, ranking, diferencial semántico.
- Distinguir **posición** de **relevancia**.
- Validar claims con MaxDiff (caso Banco X).

## PRÓXIMA CLASE

- Producción y análisis de datos.
- Software, IA, calidad de datos.
- Martes 19 de mayo · 6–9 PM

# Cierre



*Lo que aprendimos · próximo paso*

## HOY

- Operacionalizar concepto → variable.
- Tipos de preguntas: sí o no, Likert, ranking, diferencial semántico.
- Distinguir **posición** de **relevancia**.
- Validar claims con MaxDiff (caso Banco X).
- Captar preferencias indirectamente: vignette, list, conjoint, Max-Diff.

## PRÓXIMA CLASE

- Producción y análisis de datos.
- Software, IA, calidad de datos.
- Martes 19 de mayo · 6–9 PM