

FORMACIÓN PROFESIONAL · INVESTIGACIÓN COMERCIAL · UNIDAD 7

Muestreo y diseño de selección

Lunes 5 de abril de 2027. El cuestionario del Estudio de kits está probado y Nora quiere poner a Adrián en la puerta «y que pregunte a todo el que salga». ¿A quién se quiere conocer, de qué lista se saca la muestra y cómo se elige cada hogar para poder generalizar?

PARA EMPEZAR

En esta unidad aprenderás a...



Definir unidad de muestreo, población objetivo y marco, y revisar la cobertura y los duplicados del marco.



Seleccionar un aleatorio simple sin reposición y un sistemático con arranque aleatorio, de forma reproducible.



Distinguir estratos, conglomerados, áreas, etapas y rutas, y diseñar el esquema adecuado a una población.



Aplicar conveniencia, juicio, cuotas y bola de nieve, y explicar qué permiten concluir.



Documentar la selección en un protocolo y comparar métodos por coste y límites.

SITUACIÓN DE PARTIDA

«Que pregunte a todo el que salga»

Lunes 5 de abril de 2027. El cuestionario del Estudio de kits de **La Despensa de Nora** (caso ficticio) está probado y toca decidir a quién se le pasa. Nora lo tiene claro: «El viernes por la mañana ponemos a Adrián en la puerta y que pregunte a todo el que salga». Unai añade: «Y yo lo cuelgo en las redes».

Tu equipo de análisis tiene que contestar **antes del viernes 9**, con el protocolo de selección por escrito: a quién se quiere conocer (hogares y oficinas que podrían comprar kits, no solo quien ya entra en la tienda), de qué lista se puede sacar la muestra y qué falta en esa lista, y cómo se elige cada hogar para poder generalizar, y hasta dónde.

❓ **¿Qué debe decidir Nora: cuántas respuestas quiere o a quién quiere poder describir con ellas?**

❓ **¿Qué sabemos de quien sale de la tienda una mañana, y qué suponemos de quien no sale?**

❓ **Si preguntas a cien personas en la puerta, ¿a qué población representan esas cien respuestas?**

❓ **¿Qué prueba permitiría revisar la primera respuesta de Nora antes de salir a la calle?**

MAPA DE LA UNIDAD

Lo que vamos a ver

7.1 Población y marco

7.2 Aleatorio simple y sistemático

7.3 Estratificado y conglomerados

7.4 Conveniencia, cuotas y bola de nieve

7.5 Sesgos y selección documentada

¿A quién dejaremos fuera?



Preguntemos a los primeros que pasen.

¿A quién dejaremos fuera? 🤔



Definamos población y selección.

CLAVE

Antes de calcular cuántos, hay que saber **quiénes**. Una muestra perfecta sacada de la lista equivocada responde a otra pregunta.

Quiénes, qué se sortea y de qué lista



Población objetivo

Unidades sobre las que la decisión necesita información, por **contenido, ámbito y tiempo**: hogares del barrio y oficinas del barrio y del polígono, en abril de 2027. También se llama universo.



Muestreo, análisis e informante

Se sortea el hogar (unidad de muestreo), se obtienen datos del hogar (análisis) y responde quien hace la compra (informante). En oficinas, la empresa; contesta quien propone las compras, como Mencía en Albada Asesores.



Marco muestral

La lista, el mapa o el registro del que se seleccionan las unidades. Solo lo que está en el marco puede salir.



Población de estudio

La que de verdad se puede describir: la que define el marco. Sortear tiques en lugar de hogares daría más peso a quien compra más veces.

CLAVE

Antes de sortear se revisan la **subcobertura** (unidades de la población que no están en el marco), la **sobrecobertura** (unidades del marco que no son de la población) y los **duplicados** (más probabilidad de salir).

7.1 · TABLA 1

Población objetivo y marcos posibles del Estudio de kits

MARCO POSIBLE	UNIDAD	QUÉ DEJA FUERA	USO
Personas que salen de la tienda	Persona	No clientes, oficinas, quien compra por la web	No es un marco: nadie lo puede numerar
Titulares de la tarjeta de cliente	Titular	No clientes; un hogar puede tener dos tarjetas	Clientes con tarjeta
Lista de interés de kits (con autorización)	Hogar	Hogares que no han pedido información	Prueba con hogares interesados
Directorio de empresas del polígono	Empresa	Oficinas del barrio; autónomos sin local	Oficinas del polígono

La primera fila es la propuesta de Nora: no hay lista y quien nunca entra no tiene ninguna posibilidad de salir. La tarjeta de cliente (350 titulares el año pasado) describe a la clientela con tarjeta, no a todos los que compran en el supermercado.

CLAVE

El marco elegido decide a quién se podrá generalizar: la lista de interés describe a los **hogares interesados**, no al barrio.

Siempre el mismo orden, y cada fase deja registro

1 → 2 → 3 → 4 → 5 → 6

1 · Población y unidad
Definir la población objetivo y la unidad de muestreo.

2 · Marco
Elegir o construir el marco y depurarlo.

3 · Método
Elegir el método de selección (7.2-7.4).

4 · Tamaño
Determinar el tamaño de la muestra (unidad 8).

5 · Selección y campo
Seleccionar las unidades y hacer el trabajo de campo con registro de contactos.

6 · Revisión
Revisar cobertura, sustituciones y no respuesta, y documentarlo (7.5).

CLAVE
El tamaño va **después** de la población, el marco y el método: no se decide cuántos antes de saber quiénes y cómo.

7.1 · REGISTRO 1

Depuración de la lista de interés de kits

Martes 6 de abril: Adrián revisa la lista de hogares que pidieron información de los kits entre el 11 de enero y el 31 de marzo y autorizaron que se les contactara para estudios.

REVISIÓN	FILAS	DECISIÓN
Filas de la lista original	104	—
Mismo hogar con dos correos (mismo teléfono y portal)	3	Se deja la primera alta
Una oficina del barrio apuntada como hogar	1	Fuera: es de otra población
Marco depurado (H001 a H100, por fecha de alta)	**100**	Versión 1.0, 06/04/2027

Los 3 duplicados son un problema de duplicados; la oficina, de sobrecobertura. El archivo original se conserva sin tocar y se trabaja sobre una copia con versión.

CLAVE

Control: **104 − 3 − 1 = 100** hogares, numerados H001 a H100 por fecha de alta.

Minimización de datos

⚖️

Los datos personales serán «adecuados, pertinentes y limitados a lo necesario en relación con los fines para los que son tratados («minimización de datos»)».

Reglamento (UE) 2016/679 (Reglamento general de protección de datos), de 27 de abril de 2016, art. 5.1 c) · Consulta la versión vigente

SUJETO OBLIGADO

Quien trata los datos del marco: para sortear basta código, zona y contacto autorizado. La lista de interés solo incluye hogares que autorizaron que se les contactara para estudios.

CLAVE

El marco depurado trabaja con identificadores (H001 a H100), no con nombres: es lo que necesita el sorteo.

🙋 PREGUNTA PARA LA CLASE

La lista de interés no contiene a los hogares que nunca pidieron información. Es un problema de…

A Sobrecobertura del marco

B Subcobertura del marco

C Duplicados del marco

✔ Subcobertura: unidades de la población que no están en el marco. La oficina apuntada como hogar sería sobrecobertura, y el hogar con dos correos, un duplicado (autoevaluación 1).

Que decida un procedimiento, no una persona

🎯

Probabilidad de inclusión

La que tiene cada unidad del marco de entrar en la muestra. Si se conoce para todas y es mayor que cero, se puede estimar con un margen de error; si no, la muestra sirve para explorar.

🎲

Aleatorio simple sin reposición

Todas las muestras posibles de tamaño n tienen la misma probabilidad. Con N = 100 y n = 10: probabilidad **10/100 = 0,10** y **17.310.309.456.440** muestras distintas.



Tres pasos en la hoja

Columna A, H001 a H100; columna B, =ALEATORIO(); pegar **solo los valores**; ordenar por B, tomar las n primeras y guardar con fecha, versión y autor.

🔑

Semilla

Con el mismo generador, semilla y marco sale la misma muestra: anótala. Ejemplo ilustrativo (Python, semilla 2027, marco v1.0): H002, H011, H014, H024, H042, H045, H052, H060, H081 y H084. No uses =ALEATORIO.ENTRE(1;100) diez veces: puede repetir un número (reposición).

CLAVE

Aleatorio no es «al azar». Preguntar «a cualquiera» no es aleatorio: decide quien pregunta. Aleatorio es un procedimiento escrito que no depende de nadie.

Un intervalo, un arranque sorteado

k = N / n · a, a + k, a + 2k…

Intervalo	100 / 10	**k = 10**
Arranque sorteado entre 1 y k	=ALEATORIO.ENTRE(1;10)	**a = 7**
Selección	7, 17, 27… 97	**10 hogares**
Probabilidad de inclusión	1 / k = 1 / 10	**0,10**
Muestras posibles	una por arranque	**solo k = 10**

Es rápido y reparte la muestra a lo largo de la lista. Pero dos unidades vecinas nunca pueden salir juntas: por eso no es un aleatorio simple, aunque tenga la misma probabilidad de inclusión.

CLAVE

Periodicidad: si la lista tiene un ciclo que coincide con k o con un divisor de k, falla. Ordenada por portales de diez viviendas, con k = 10 y arranque 7 saldría siempre la séptima vivienda de cada portal.

Se parecen en el dibujo, son opuestos en la lógica



Estratificar: dividir para que no falte nadie

Grupos que no se solapan, cubren todo el marco y son **parecidos por dentro**. Se sortea dentro de **cada** estrato. Hay que conocer el estrato de cada unidad antes de sortear (zona, tipo de hogar u oficina, antigüedad como cliente).

VS



Conglomerar: agrupar para poder hacer el campo

Grupos naturales (una empresa con su plantilla, una manzana con sus viviendas) que se sortean **enteros**. Conviene que sean **variados por dentro**. Ejemplo: de 30 empresas del polígono se sortean 5.

CLAVE

Si sus unidades se parecen mucho, cada entrevista de un conglomerado aporta menos y hace falta más muestra.

7.3 · EJEMPLO DEL LIBRO

Una muestra de 10 proporcional a las zonas

En el marco depurado, cada hogar lleva su zona: Norte 40, Centro 36 y Sur 24.

ZONA	HOGARES	CUOTA EXACTA	PARTE ENTERA	RESTO	MUESTRA
Norte	40	4,0	4	0,0	4
Centro	36	3,6	3	0,6	4
Sur	24	2,4	2	0,4	2
Total	**100**	**10,0**	**9**		**10**

Las partes enteras suman 9: el hogar que falta va al estrato con mayor resto (Centro, 0,6). El cálculo del reparto entre estratos se trabaja a fondo en la unidad 8.

CLAVE

Estratificar **garantiza que ningún grupo falte** y gana precisión si los estratos son distintos entre sí.

7.3 · CUADRO 1

Seis diseños probabilísticos comparados

DISEÑO	QUÉ SE SORTEA	EJEMPLO EN LA DESPENSA	LÍMITE PRINCIPAL
Aleatorio simple	Unidades del marco	10 hogares de H001-H100	Marco completo; muestra dispersa
Sistemático	Arranque; luego cada k	7, 17… 97 del marco	Ciclos del marco
Estratificado	Unidades dentro de cada estrato	Hogares por zona	Estrato conocido para todos
Conglomerados	Grupos enteros	5 de 30 empresas	Grupos parecidos por dentro
Por áreas y etapas	Zonas, luego viviendas, luego persona	Manzanas del barrio	Cálculo más complejo
Ruta aleatoria	Punto de partida y reglas	Portales desde un cruce sorteado	Solo con reglas documentadas

Polietápico: cada etapa tiene su marco y su sorteo, y la probabilidad final es el producto de las de cada etapa. El INE describe su TIC-H 2025 como un muestreo trietápico estratificado: secciones censales, viviendas familiares principales y una persona de 16 o más años (26.862 viviendas de tamaño teórico).

CLAVE

La **ruta aleatoria** solo es probabilística si puntos, hogares, persona y sustituciones están documentados; si el entrevistador puede saltarse portales, es conveniencia.

🙋 PREGUNTA PARA LA CLASE

Itxaso dice que «estratificar por zona y sortear manzanas es lo mismo». ¿Tiene razón?

A Sí: en los dos se divide el barrio en grupos

B No: al estratificar se sortea dentro de cada zona y ninguna falta; al sortear manzanas se eligen grupos enteros y algunas zonas pueden quedar fuera

C No: sortear manzanas no es probabilístico

✔ Estratificar es dividir para que no falte ningún grupo; conglomerar es agrupar para que el campo sea posible. Para que no falte ninguna zona, se estratifica (actividad 6).

Rápidos, a veces los únicos, sin margen de error



Conveniencia

Se pregunta a quien está a mano. La propuesta de Nora deja fuera a quien trabaja a esa hora, a las oficinas, a quien compra solo por la web y a quien nunca entra.



Juicio (u opinática)

Una experta elige a quienes considera típicos: Olvido y cinco habituales «que saben de cocina». Útil para un pretest; refleja lo que Olvido cree típico.



Cuotas

Se fija cuántas personas de cada grupo y el entrevistador **elige libremente** dentro de cada cuota. Unai: 10 hogares y 10 oficinas; en cada grupo, 5 de menos de 45 años y 5 de 45 o más.



Bola de nieve

Cada participante propone a otros del perfil: Mencía presentaría a otras responsables de servicios generales. Riesgo: un círculo de conocidos parecidos.

CLAVE

Cuotas cumplidas no es lo mismo que muestra probabilística: controlan la composición en unas pocas variables, no quién entra. Sirven para explorar motivos y lenguaje, probar un instrumento o encontrar casos raros.

7.5 · TABLA 2

Sesgos de selección más frecuentes

SESGO	QUÉ OCURRE	EJEMPLO
Cobertura	El marco no contiene a parte de la población	La lista de interés no tiene no clientes
Selección	El procedimiento favorece a unos	Preguntar en la puerta por la mañana
Autoselección	Responde quien quiere	Encuesta abierta en las redes
No respuesta	Quien no contesta es distinto	Hogares que trabajan hasta tarde
Sustitución	El entrevistador cambia a quien no encuentra	Pasar al vecino que abre la puerta

Grande no es representativo: en marzo, una encuesta abierta en la web reunió 286 respuestas; 241 eran de personas que ya compraban en la web (84,3 %). Describen a la clientela digital, no al barrio.

CLAVE

El **error de muestreo** se calcula y baja al aumentar la muestra; un **sesgo** empuja siempre hacia el mismo lado y **no** baja con más muestra.

7.5 · DOCUMENTO 1

Protocolo de selección EK-2027/PS-v1.0

CAMPO	CONTENIDO
Población de estudio	Hogares de la lista de interés de kits (11/01-31/03/2027)
Marco	Lista depurada v1.0 (06/04/2027): 100 hogares, H001-H100
Unidad e informante	Hogar; persona que hace la compra
Método	Sistemático, k = 10, arranque sorteado el 08/04
Sustituciones	No se sustituye; se registra el motivo
Registro de contactos	Fecha, hora, intento (máximo 3) y resultado
Responsables	Diseño: equipo de análisis · registro: Adrián · aprueba: Nora

Si un hogar no contesta: los intentos previstos en horarios distintos, cada resultado registrado (completa, rechazo, no localizado, fuera del marco) y, si se sustituye, con una reserva sorteada a la vez que la muestra y en el orden sorteado.

CLAVE

El diseño se escribe **antes** del campo y lo que pasa **durante** el campo se registra. Lo que no está escrito no se puede revisar.

CASO RESUELTO · PASO A PASO

El sistemático de la lista de interés



El equipo de análisis, con Adrián

Jueves 8/04/2027: 10 de los 100 hogares del marco depurado para una entrevista telefónica sobre la entrega programada de kits. Caso ficticio.

1. **Decisión, datos y alcance.** Qué hogares se llaman, sin que nadie los elija a mano. Marco v1.0, H001-H100 por fecha de alta, sin duplicados; n = 10. Describirá a los hogares de la lista, no al barrio.

2. **Procedimiento.** k = 100/10 = 10. Arranque con =ALEATORIO.ENTRE(1;10), pegado como valor y anotado: 7. Selección: 7, 17, 27… 97 (107 ya está fuera del marco).

3. **Comprobación.** 10 identificadores distintos dentro del marco; altas por fecha, de una a seis al día; la zona de alta no se repite cada 10 posiciones. Zonas: 4 Norte, 4 Centro y 2 Sur (marco 40, 36 y 24).

4. **Qué no permite.** No es un aleatorio simple (solo 10 muestras posibles); con 10 hogares cualquier porcentaje será muy impreciso; nada se puede decir de quien no está en la lista.

✔ **H007, H017, H027, H037, H047, H057, H067, H077, H087 y H097**, cada hogar con probabilidad 0,10; conclusiones válidas solo para los 100 hogares de la lista de interés.

🙋 PREGUNTA PARA LA CLASE

Ruedalia: 60 clientes numerados por fecha de alta, 6 llamadas, «uno de cada diez» y en el sorteo entre 1 y 10 sale el 4. ¿Qué es correcto?

A Es un aleatorio simple porque el 4 salió por sorteo

B La selección es 4, 14, 24, 34, 44 y 54, y es un sistemático

C El intervalo es 6

✔ k = 60/6 = 10; arranque 4. Cada cliente tiene probabilidad 1/10, pero solo hay 10 muestras posibles: es un sistemático con arranque aleatorio, no un aleatorio simple (autoevaluación 7).

APRENDE CON LA IA

La respuesta que recibió Unai

Unai prepara sus 20 entrevistas con cuotas y pregunta a un asistente de IA si puede dar un margen de error. Pulsa cada tarjeta.

«Cumplir cuotas de edad convierte automáticamente la muestra en probabilística»

BULO

Dentro de cada cuota elige el entrevistador: la probabilidad de inclusión no se conoce (7.4).

«Con 20 entrevistas puedes usar la fórmula del aleatorio simple»

BULO

La fórmula exige un sorteo del marco. Sin probabilidad de inclusión conocida no hay margen de error (7.2).

«El resultado vale para todo el barrio»

BULO

En la puerta faltan quienes nunca vienen y las oficinas por la mañana: se informa como «de quienes respondieron».

«Si alguien no quiere contestar, sustitúyelo por el siguiente que pase»

BULO

Es un sesgo de sustitución (tabla 2): la muestra se desplaza hacia quien es fácil de encontrar.

El protocolo de selección

Para el negocio de tu investigación, con datos ficticios o propios autorizados:

1

Población y marco

Población objetivo y de estudio; marco con versión, fecha y registro de depuración (filas, duplicados, ajenos).

2

Unidad y método

Unidad de muestreo, de análisis e informante; método elegido y alternativa descartada, con su motivo.

3

Selección

Hecha con su arranque o semilla y el archivo guardado; reglas de sustitución y registro de contactos.

4

Responsable

Responsable y fecha de revisión. Versiona el protocolo: si cambias algo, conserva el motivo.

CLAVE

Comprueba tu paso: **cualquiera puede repetir tu selección** con tu archivo; ningún porcentaje lleva margen de error si la selección no es probabilística.

REPASO

Ideas clave

- La población objetivo se define por contenido, ámbito y tiempo; el marco decide la población de estudio, la que de verdad se puede describir.
- Unidad de muestreo (lo que se sortea), de análisis (de qué se obtienen datos) e informante (quien contesta).
- Antes de sortear se depura el marco: subcobertura, sobrecobertura y duplicados. Lista de interés: 104 − 3 − 1 = 100 hogares.
- Una muestra es probabilística si cada unidad tiene una probabilidad conocida y mayor que cero de salir y la elige un procedimiento.
- Aleatorio simple sin reposición: n/N = 10/100 = 0,10 y todas las muestras igual de probables; se sortea sin repetir y se guarda la semilla.
- Sistemático: k = N/n y arranque sorteado entre 1 y k. Misma probabilidad (1/k), pero solo k muestras posibles; cuidado con los ciclos de la lista.
- Estratificar es dividir para que no falte ningún grupo; conglomerar, agrupar para que el campo sea posible. Áreas, etapas y rutas, con reglas escritas.
- Conveniencia, juicio, cuotas y bola de nieve sirven para explorar: resultados de quienes respondieron, sin margen de error.
- Un tamaño grande no corrige una selección torcida: 286 respuestas en la web, el 84,3 % de clientes digitales. El protocolo se escribe antes del campo.



¡Ahora te toca a ti!

Repasa, practica y comprueba lo que has aprendido.

📝 **Resumen**

✅ **Test**

👆 **Interactivos**

Libro, presentaciones, resúmenes y tests gratis en **josesande.com**