COMERCIO Y MARKETING · GRADO SUPERIOR

Investigación comercial

Módulo 1010



Jose Sande

COMPARTIENDO CONOCIMIENTO

Investigación comercial

Módulo 1010 · Grado superior

Marketing y Publicidad · Gestión de Ventas y Espacios Comerciales · Comercio y Marketing

Libro del alumno · Edición genérica · Primera edición: 2026

© Jose Sande, 2026

Edita: Compartiendo Conocimiento

La Despensa de Nora, Ruedalia, sus personajes, sus clientes, sus proveedores y los demás casos del libro son ficticios.

Licencia del PDF gratuito y condiciones de uso: permiso para compartir e imprimir.

Cómo es tu libro

Tu libro tiene doce unidades de 16 páginas en cuatro bloques: 1 · Mercado y diseño de la investigación (unidades 1 a 3), 2 · Fuentes e instrumentos (unidades 4 a 6), 3 · Muestreo y preparación de datos (unidades 7 a 9) y 4 · Análisis y comunicación de resultados (unidades 10 a 12). Todas las unidades siguen el mismo recorrido.

APERTURA	**Págs. 1-2 de cada unidad.** Una ilustración, lo que vas a aprender, la situación de partida, preguntas para empezar y el mapa de la unidad.
APARTADOS	**Págs. 3-10.** La explicación con los casos de La Despensa de Nora: tablas, cálculos completos, documentos, un cómic, cajas y actividades al final de cada apartado.
CASO RESUELTO	**Pág. 11.** Un caso de La Despensa resuelto paso a paso, con un resultado que puedes comprobar.
PRÁCTICA	**Pág. 12.** La misma tarea en Ruedalia, un taller de bicicletas, con una plantilla para completar.
LECTURA Y DEBATE	**Pág. 13.** Una historia breve con preguntas para pensar y debatir.
APRENDE CON LA IA	**Pág. 13.** La respuesta de una IA con un error: lo buscas, lo compruebas y la corriges.
ACTIVIDADES FINALES	**Pág. 14.** Comprueba, Aplica e Investiga, y el paso de «Tu investigación avanza».
OJO PROFESIONAL	**Pág. 15.** Un documento de La Despensa con errores para detectar y corregir.
AUTOEVALUACIÓN	**Pág. 16.** Diez preguntas y la lista «Ya sé...».

Las cajas

Clave. La idea esencial, una definición o un error frecuente.

La norma dice. Lo que dice la norma, con su artículo y su fecha. Consulta siempre la versión vigente.

En tu sector. Cómo se aplica en otros negocios: comercio, hostelería, servicios, industria, agencias...

Para saber más. Una ampliación para quien quiera ir más allá.

Tu investigación avanza. Un paso de tu propia investigación comercial en cada unidad.

Personajes. La Despensa de Nora es una tienda de alimentación de barrio con obrador propio que en 2027 empieza a vender kits de comida a hogares y oficinas. Nora, su dueña, tiene que decidir qué ofrecer, a quién y cómo, y te encarga a ti, como equipo de análisis, la investigación. Te ayudan Olvido, Itxaso, Unai y Adrián. Los conocerás en la página 8.

Índice

BLOQUE 1 · MERCADO Y DISEÑO DE LA INVESTIGACIÓN

Unidad 1 · Mercado, entorno y consumidor 9
1.1 Sistema comercial 11
1.2 Microentorno y macroentorno 12
1.3 Instituciones y fuentes 14
1.4 Necesidades y compra 15
1.5 Segmentación y estrategia 17
Caso resuelto, práctica, lectura y actividades finales 19

Unidad 2 · Sistema de información de marketing 25
2.1 Necesidades y decisiones 27
2.2 Fuentes y tipos de datos 28
2.3 Subsistemas del SIM 29
2.4 Calidad y periodicidad 31
2.5 Registro y consulta 33
Caso resuelto, práctica, lectura y actividades finales 35

Unidad 3 · Problema, objetivos y diseño 41
3.1 Decisión y problema 43
3.2 Objetivos e hipótesis 44
3.3 Exploratorio, descriptivo y causal 46
3.4 Método y fuentes 47
3.5 Plan, presupuesto y calendario 48
Caso resuelto, práctica, lectura y actividades finales 51

BLOQUE 2 · FUENTES E INSTRUMENTOS

Unidad 4 · Fuentes secundarias y suficiencia 57
4.1 Búsqueda online y offline ... 59
4.2 Fiabilidad y representatividad ... 61
4.3 Organización y metadatos ... 62
4.4 Suficiencia y lagunas ... 64
4.5 Archivo y recuperación ... 66
Caso resuelto, práctica, lectura y actividades finales ... 67

Unidad 5 · Investigación cualitativa 73
5.1 Entrevista en profundidad ... 75
5.2 Dinámica de grupo ... 76
5.3 Observación ... 78
5.4 Técnicas proyectivas y creatividad ... 80
5.5 Guion, registro y análisis ... 81
Caso resuelto, práctica, lectura y actividades finales ... 83

Unidad 6 · Encuestas y cuestionarios 89
6.1 Modalidades de encuesta ... 91
6.2 CAPI, MCAPI, CATI y CAWI ... 92
6.3 Redacción y escalas ... 93
6.4 Saltos y prueba piloto ... 95
6.5 Paneles y elección de técnica ... 97
Caso resuelto, práctica, lectura y actividades finales ... 99

BLOQUE 3 · MUESTREO Y PREPARACIÓN DE DATOS

Unidad 7 · Muestreo y diseño de selección 105
7.1 Población y marco ... 107
7.2 Aleatorio simple y sistemático ... 109
7.3 Estratificado y conglomerados ... 110
7.4 Conveniencia, cuotas y bola de nieve ... 112
7.5 Sesgos y selección documentada ... 113
Caso resuelto, práctica, lectura y actividades finales ... 115

Unidad 8 · Tamaño, precisión y no respuesta 121
8.1 Confianza y error 123
8.2 Variabilidad y población finita 124
8.3 Cálculo y redondeo 126
8.4 Asignación por estratos 127
8.5 No respuesta y limitaciones 129
Caso resuelto, práctica, lectura y actividades finales 131

Unidad 9 · Depuración, codificación y descripción 137
9.1 Diccionario y validación 139
9.2 Ausencias, duplicados y filtros 141
9.3 Frecuencias y gráficos 143
9.4 Tendencia central y dispersión 144
9.5 Forma de la distribución 146
Caso resuelto, práctica, lectura y actividades finales 147

BLOQUE 4 · ANÁLISIS Y COMUNICACIÓN DE RESULTADOS

Unidad 10 · Inferencia e interpretación 153
10.1 Población, muestra y estimador 155
10.2 Intervalos de confianza 156
10.3 Error muestral y sesgo 157
10.4 Comparaciones y causalidad 159
10.5 Conclusiones para decidir 161
Caso resuelto, práctica, lectura y actividades finales 163

Unidad 11 · Base de datos relacional 169
11.1 Entidades, claves y relaciones 171
11.2 Diseño e integridad 172
11.3 Consultas y búsquedas 174
11.4 Formularios y validación 176
11.5 Seguridad, mantenimiento e informes 177
Caso resuelto, práctica, lectura y actividades finales 179

Unidad 12 · Informe, archivo y defensa 185
12.1 Objetivo y método 187
12.2 Resultados y gráficos 188
12.3 Conclusiones y límites 190
12.4 Recomendación y destinatario 191
12.5 Actualización y transferencia 193
Caso resuelto, práctica, lectura y actividades finales 195

Tu investigación avanza

En la página 14 de cada unidad, «Tu investigación avanza» te pide un paso de tu propia investigación comercial sobre un negocio que elijas o te proponga el profesorado: de la pregunta de decisión al informe final. Cada paso es una entrega con su código, de I01 a I12.

Los pasos de tu investigación

Paso 1. I01 · Mercado y pregunta de decisión (unidad 1) 22
Paso 2. I02 · Diseño del SIM (unidad 2) 38
Paso 3. I03 · Plan de investigación (unidad 3) 54
Paso 4. I04 · Inventario y lagunas (unidad 4) 70
Paso 5. I05 · Guion cualitativo y codificación (unidad 5) 86
Paso 6. I06 · Cuestionario probado (unidad 6) 102
Paso 7. I07 · Protocolo de selección (unidad 7) 118
Paso 8. I08 · Tamaño y contactos (unidad 8) 134
Paso 9. I09 · Datos depurados y descripción (unidad 9) 150
Paso 10. I10 · Inferencia y conclusión (unidad 10) 166
Paso 11. I11 · Base relacional comprobada (unidad 11) 182
Paso 12. I12 · Informe y transferencia al SIM (unidad 12) 198

Cómo trabajarás. Cada paso lleva versión, fecha, la fuente de cada dato y una comprobación, y otra persona lo revisa. Una cifra necesita su unidad, su base, su fuente, su fecha y su límite. Trabaja con hoja de cálculo, formulario y gestor de bases de datos.

Mi registro de la investigación

Paso 1.	Versión _____	Fecha _____	Revisado por _____
Paso 2.	Versión _____	Fecha _____	Revisado por _____
Paso 3.	Versión _____	Fecha _____	Revisado por _____
Paso 4.	Versión _____	Fecha _____	Revisado por _____
Paso 5.	Versión _____	Fecha _____	Revisado por _____
Paso 6.	Versión _____	Fecha _____	Revisado por _____
Paso 7.	Versión _____	Fecha _____	Revisado por _____
Paso 8.	Versión _____	Fecha _____	Revisado por _____
Paso 9.	Versión _____	Fecha _____	Revisado por _____
Paso 10.	Versión _____	Fecha _____	Revisado por _____
Paso 11.	Versión _____	Fecha _____	Revisado por _____
Paso 12.	Versión _____	Fecha _____	Revisado por _____

La Despensa de Nora, S.L.

Los casos transcurren en 2027 en La Despensa de Nora, S.L., una tienda de alimentación de proximidad de un barrio de una ciudad media, con obrador propio, tienda en línea y, desde este año, una línea de kits de comida para hogares y oficinas. A 300 metros abrió hace un año un supermercado de cadena. La Despensa, sus clientes, sus proveedores, Ruedalia y las cifras de los casos son ficticios.



Dirección. Nora Castro, socia única y gerente: decide y firma.

Obrador. Olvido, panadera: conoce el producto y a la clientela de siempre.

Kits y pedidos. Itxaso: fichas, preparación, capacidad y entregas a hogares y oficinas.

Comunicación digital. Unai: web, redes, correo y campañas.

Registros y comprobaciones. Adrián, alumno de grado medio en formación en empresa de enero a junio.

Quién es quién

Nora (48 años), dueña de La Despensa. Quiere decidir bien y deprisa; a veces busca datos que le den la razón.

Olvido (59 años), panadera del obrador. Conoce a cada cliente habitual, pero no a los que nunca entran.

Itxaso (38 años), jefa de kits y pedidos. «Si no cabe en la mañana, no lo prometas».

Unai (25 años), comunicación digital. Propone mucho y rápido: una encuesta web, preguntar a quien pasa…

Adrián (20 años), en formación en empresa de enero a junio. Cuenta, anota, compara y detecta los datos raros.

Tú, el equipo de análisis. Nora te encarga cada paso de la investigación.



UNIDAD 1

Mercado, entorno y consumidor

Bloque 1 · Mercado y diseño de la investigación

EN ESTA UNIDAD APRENDERÁS A…

- Delimitar un mercado y un problema de decisión antes de medir nada, y separar las variables que la empresa controla de las que no controla.
- Construir un mapa del microentorno y del macroentorno con relaciones, efectos sobre la decisión, indicadores y fuentes.
- Identificar instituciones y fuentes por su función y por el tipo de dato que aportan, y juzgar si responden a tu pregunta.
- Comparar necesidades y procesos de compra de un hogar y de una organización, con sus decisores e influencias.
- Segmentar registros con criterios operativos y elegir una estrategia justificada por recursos y necesidades.

SITUACIÓN DE PARTIDA

Lunes 11 de enero de 2027. La Despensa de Nora, S.L. estrena su línea de kits de comida para hogares y oficinas. En la trastienda, Nora reúne a su gente. Olvido lo tiene claro: «Los habituales me dicen que les gusta». Nora no lo discute, pero tiene otra pregunta: el supermercado cercano sigue lleno a las siete de la tarde y en el polígono hay oficinas que no han pisado nunca la tienda. ¿Por qué otros no entran?

Nora tiene que decidir, antes de marzo, si los kits se orientan a familias del barrio, a oficinas o a ambos. Te encarga a ti, como equipo de análisis, que ordenes lo que se sabe y lo que se supone y que conviertas su duda en una **pregunta de decisión** que se pueda medir. La quiere el viernes 22 de enero.

PARA EMPEZAR

1. ¿Qué tiene que decidir Nora exactamente: qué kit, para quién, a qué precio, con qué entrega? ¿Qué parte de esa decisión depende solo de ella?
2. ¿Qué sabe La Despensa por sus clientes habituales y qué está suponiendo sobre quienes no entran?
3. ¿Qué dato o prueba permitiría revisar la primera respuesta de Olvido antes de convertirla en una decisión?



Sistema comercial

1.1



Micro y macro

1.2



Instituciones y fuentes

1.3



Necesidades y compra

1.4



Segmentación

1.5

1.1 Sistema comercial

Antes de preguntar nada a nadie hay que saber qué se va a decidir, sobre qué mercado y con qué margen de maniobra. Una categoría mal definida hace que ningún dato sea comparable.

De la tienda al sistema comercial. El **sistema comercial** es el conjunto formado por la empresa, el mercado al que se dirige y el entorno en el que ambos se mueven. Una **decisión empresarial** es una elección entre alternativas que la empresa puede ejecutar: lanzar o no un kit, fijar su precio, abrir un horario de entrega. La investigación comercial no decide por la empresa: reduce la incertidumbre de esa elección.

Variables controlables y no controlables. Las **controlables** son las que la empresa fija por sí misma: las del **marketing mix** (producto, precio, distribución y comunicación) y sus recursos (capacidad de preparación, plantilla, horarios). Las **no controlables** vienen de fuera: competidores, renta de los hogares, normas, tecnología. La empresa no puede cambiarlas, pero sí **anticiparse** y **adaptarse**. La conducta de los clientes (a qué hora comen, cuándo salen de trabajar) tampoco la controla: es un **dato que hay que medir**, no una palanca.

CLAVE · INFLUIR NO ES CONTROLAR

La Despensa puede influir en una oficina con una buena oferta, pero no controla su presupuesto ni su hora de comer. Pregúntate: ¿puede la empresa cambiar este factor mañana, con una decisión propia? Si no, es no controlable, por mucho que influya en él.

Estructura y límites del mercado. Un **mercado** es el conjunto de compradores, actuales y potenciales, de un producto, y se clasifica con varios criterios a la vez:

Tabla 1 · Clasificación del mercado de los kits

Criterio	Tipos	Los kits de La Despensa
Comprador	Consumo u organizacional	Hogares y oficinas: los dos
Producto	Bienes de consumo o duraderos; servicios	Consumo inmediato con entrega
Ámbito geográfico	Local, regional, nacional, internacional	Local: barrio y polígono
Competencia	Monopolio, oligopolio, monopolística, perfecta	Monopolística, a verificar
Desarrollo	Actual, potencial, tendencial	Actual pequeño, nace en 2027

Delimitar el mercado y el problema. Antes de medir hay que fijar los **límites** del mercado: qué producto o necesidad, qué clientes, qué territorio y qué periodo. «Comida para oficinas» es una categoría mal definida: mezcla el menú del bar, el táper de casa y el catering de una reunión, y dos personas que cuenten cosas distintas obtienen cifras incomparables. En el caso: **comida preparada para recoger o recibir, para hogares del barrio y oficinas del barrio y del polígono, en días laborables de 2027**.

Con el mercado delimitado se formula el **problema de decisión**, que no es el **síntoma**. «Entran menos familias por la tarde» es un síntoma; «decidir si los kits se dirigen a hogares, a oficinas o a ambos, y con qué servicio de pedido» es el problema. Un buen problema nombra la decisión, quién la toma y para cuándo; si mezcla síntoma, causa supuesta y solución deseada («no vienen porque el súper es más barato: bajemos precios»), la investigación nace sesgada.

Ejemplo resuelto. Del comentario de Olvido al problema. Dato: varios habituales han comentado en el mostrador que les gustan los kits (fuente: Olvido; diciembre y enero; sin registro escrito). **Interpretación:** «a los clientes les gustan». **Lo que no dice:** si comprarían, cuántos, a qué precio, ni qué piensan quienes no son clientes. **Problema:** «decidir, antes del 1 de marzo de 2027, a qué grupo dirige La Despensa sus kits y con qué forma de entrega».

ACTIVIDADES

1. Clasifica como variable controlable o no controlable el precio propio, la capacidad de la zona de kits, el supermercado a 300 m y la renta del barrio. Para cada uno, explica qué puede hacer La Despensa (cambiarlo, adaptarse o vigilarlo) y en qué plazo.
2. Delimita el mercado del K03 Comida de trabajo (necesidad, clientes, territorio y periodo) y clasifícalo con los cinco criterios de la tabla 1.
3. Indica si cada frase es un síntoma, una causa supuesta o un problema de decisión: a) «Las oficinas del polígono no nos conocen»; b) «Este mes hemos vendido pocos kits K04»; c) «Decidir antes de marzo si abrimos la entrega programada a oficinas».

1.2 Microentorno y macroentorno

Unos agentes tratan a diario con la tienda y otras fuerzas cambian las reglas para todos. Hay que saber cuáles son, cómo afectan a la decisión y con qué dato se siguen.

Cercanía y relación directa. El **microentorno** son los agentes cercanos con **relación directa** con la empresa: clientes, proveedores, competidores, intermediarios y públicos (vecinos, asociación de comerciantes, medios locales). El **macroentorno** son las fuerzas generales que afectan a todas las empresas del mercado, en seis grupos: **económico** (renta, precios, empleo), **demográfico** (edad, tamaño de los hogares), **sociocultural** (hábitos de comida, teletrabajo), **tecnológico** (pedidos en línea, pago con el móvil), **medioambiental** (envases, residuos) y **político-legal** (horarios, etiquetado, datos personales). Micro y macro **no** miden el tamaño del agente ni la importancia del factor: un competidor pequeño es microentorno y puede pesar más en la decisión que una tendencia nacional.

Del listado al mapa. No basta con enumerar. Un **mapa del entorno** dice, para cada factor, su nivel, si la empresa lo controla, **qué efecto concreto** tiene sobre la decisión y **con qué indicador y fuente** se sigue. Es el recurso técnico de esta unidad. Un mismo agente puede tener dos papeles: las **plataformas de kits a domicilio** son intermediarias (llevan pedidos de otros) y competidoras a la vez, y hay que decir qué papel pesa en cada decisión.

Tabla 2 · Mapa de variables controlables, micro y macro de los kits

(borrador del equipo, 18/01/2027)

Factor y nivel	Efecto en la decisión	Indicador y fuente
Precio y kits (empresa, controlable)	Define el público al que llega	Ventas por kit (caja)
Zona de kits (empresa, controlable)	Limita pedidos por turno	Kits por turno (Itxaso)
Hogares y oficinas (micro: clientes)	Hogares de tarde, oficinas a mediodía	Pedidos por franja (web)
Huertas del Ribazo, Envaserde (micro: proveedores)	Coste y plazo del kit	Precio y plazo de entrega
Supermercado y Cocinas Brañas (micro: competidores)	Precio de referencia y hábito	Visita mensual de precios
Plataformas de kits (micro: intermediarios y competidores)	Alcance y coste de la entrega	Zonas y tiempos de entrega
Renta de los hogares (macro: económico)	Precio que se puede pedir	Renta media por hogar (INE)
Horarios comerciales (macro: político-legal)	Franjas que la tienda puede abrir	Norma vigente (BOE)

Las **oficinas** no son macroentorno por ser muchas: son clientes potenciales. Y la **renta** no se mide «a ojo»: tiene un indicador oficial con fuente, año y ámbito. Un factor macro solo sirve con un **indicador** que tenga **ámbito**, **fecha** y **fuente**: «la gente cocina menos» es una tendencia; «porcentaje de ocupados que teletrabajan algún día, encuesta oficial, año y ámbito» es un indicador. No confundas tendencia con hecho interno: reformar la zona de kits es una decisión de La Despensa. Tampoco confundas un factor con su efecto: «el teletrabajo» es el factor; «menos personas comen en la oficina los viernes» es su efecto posible, que habrá que comprobar.

Ejemplo resuelto. Un factor político-legal con su efecto. La Despensa tiene unos 110 m² de sala y vende sobre todo pan y platos preparados: por los dos motivos del art. 5 de la Ley 1/2004 puede fijar libremente sus días y horas de apertura. Efecto en la decisión: podría abrir la recogida de kits más tarde si los hogares H1 lo piden, sin esperar a un cambio normativo. Si el súper pertenece a un grupo no pyme, su horario depende de otras reglas. Indicador: «franjas de apertura permitidas», fuente BOE y boletín autonómico, con fecha de consulta.

LA NORMA DICE

Ley 1/2004, de 21 de diciembre, de Horarios Comerciales, art. 5.1: los establecimientos dedicados principalmente a la venta de «pan, platos preparados» y otros productos «tendrán plena libertad para determinar los días y horas en que permanecerán abiertos al público». Art. 5.2: la misma libertad tienen los de «superficie útil para la exposición y venta al público inferior a 300 metros cuadrados», salvo los de empresas o grupos de distribución que no sean pymes. Consulta la versión vigente.

ACTIVIDADES

4. Construye el mapa de clientes, proveedores, competidores e intermediarios de los kits. Para cada agente, escribe la relación con La Despensa y un efecto concreto sobre la decisión de Nora (no basta con nombrarlo).
5. Selecciona tres de los seis factores del macroentorno que afecten a la decisión. Para cada uno propón un indicador con su ámbito, su periodo y la fuente donde lo buscarías.
6. Unai escribe: «Cocinas Brañas sirve a todo el polígono, así que es macroentorno». Corrige la frase y explica qué criterio separa micro y macro.

1.3 Instituciones y fuentes

Una institución pública puede dar datos excelentes que no contestan tu pregunta. Primero se identifica qué hace cada una; después se decide si sirve.

Instituciones por su función. Unas instituciones **regulan** o vigilan los mercados, otras **producen** información y otras **representan** o apoyan a empresas y consumidores. Se clasifican por su **función** y por el **tipo de dato** que aportan, no por su prestigio:

Tabla 3 · Instituciones que regulan o informan sobre los mercados

Institución	Función	Qué aporta al caso
CNMC	Defiende la competencia efectiva en los mercados	Reglas de competencia, informes
Banco de España (en el Eurosistema)	Estabilidad de precios y política monetaria	Tipos de interés, precios
AEPD	Autoridad de protección de datos	Guías para encuestas y bases
INE	Estadísticas oficiales del Estado	Población, renta, hogares
CIS	Estudios de opinión periódicos	Actitudes generales, no del barrio
Consumo (ministerio, OMIC, asociaciones)	Información y defensa del consumidor	Reclamaciones, derechos, guías
Cámara de comercio	Corporación de derecho público	Censo público de empresas

Qué dice la norma de cada una. La **Comisión Nacional de los Mercados y la Competencia (CNMC)** tiene por objeto garantizar «una competencia efectiva en todos los mercados y sectores productivos, en beneficio de los consumidores y usuarios» (Ley 3/2013, art. 1.2). El **Banco de España** participa en la política monetaria del Sistema Europeo de Bancos Centrales, cuyo «objetivo principal» es «mantener la estabilidad de precios» (Ley 13/1994, art. 7). La **AEPD** es «una autoridad administrativa independiente de ámbito estatal» (Ley Orgánica 3/2018, art. 44.1). El **INE** tiene encomendada, entre otras, «la formación de los censos generales» y la publicación de sus resultados y su metodología (Ley 12/1989, art. 26). Las **cámaras de comercio** gestionan «un censo público de todas las empresas» de su demarcación (Ley 4/2014, art. 5.1.g). El **CIS** publica barómetros y estudios de opinión con su ficha técnica. Consulta la versión vigente de cada norma.

PARA SABER MÁS

El INE publica el **Atlas de distribución de renta de los hogares**, con renta media por persona y por hogar por municipio, distrito y **sección censal**. La edición publicada el 19/10/2026 da datos de 2024 (consulta del 28/10/2026).

Fiable no es lo mismo que pertinente. Una fuente oficial puede ser **fiable** (método público, definiciones estables) y no **responder** a tu pregunta. La renta por sección describe el poder de compra del barrio en 2024; no dice si esos hogares comprarían un kit en 2027 ni a qué hora. El censo de la cámara dice cuántas empresas hay, no quién decide la comida. Eso pedirá investigación propia (unidades 3 a 6).

CLAVE · ANTES DE USAR UNA FUENTE, CUATRO PREGUNTAS

¿Qué mide (personas, hogares, empresas, euros)? ¿De qué ámbito y fecha es? ¿Quién la produce y con qué método? ¿Responde a mi pregunta o solo describe el contexto?

ACTIVIDADES

7. Localiza un organismo estadístico, una entidad de consumo y una institución comercial (cámara o asociación). Para cada una, anota su web oficial, su función y qué información aportaría al estudio de kits. No le atribuyas funciones que no hayas comprobado.
8. Para cada fuente, di qué parte de la pregunta de Nora responde y cuál no: a) renta media por hogar de las secciones del barrio; b) censo de empresas de la cámara; c) un barómetro nacional del CIS sobre hábitos de alimentación.

1.4 Necesidades y compra

Un hogar y una oficina pueden pedir el mismo kit por razones distintas y decidirlo de formas muy diferentes.

Tipos de consumidor y de necesidad. Distingue al **consumidor final**, que compra para sí o para su hogar, del **consumidor organizacional** (empresa o institución), que compra para su actividad; y, por su relación con la tienda, al **habitual**, al **ocasional** y al **potencial**, que aún no compra. Las necesidades se clasifican con la **jerarquía de Maslow**: **fisiológicas** (comer), **seguridad** (un kit sin gluten fiable), **pertenencia** (comer con el equipo), **estima** (quedar bien en una reunión) y **autorrealización**. La **crítica** al modelo: las necesidades no se satisfacen en un orden fijo, conviven varias a la vez y su peso cambia con la cultura y la situación. Úsala para **ordenar prioridades con evidencia** en cada caso, no para predecir conductas: para una persona celíaca, la seguridad del K05 pasa por delante del precio; para Albada, la puntualidad de la entrega pesa más que el sabor el día de una reunión.

El proceso de compra del consumidor. Etapas: **reconocimiento** de la necesidad, **búsqueda** de información, **evaluación** de alternativas, **compra** y **experiencia posterior**. Influyen variables **internas** (motivación, percepción, actitudes, experiencia) y **externas** (familia, grupos, cultura, situación).

La compra industrial u organizacional. Tiene más fases: 1) reconocimiento del problema; 2) descripción de la necesidad; 3) especificación del producto; 4) búsqueda de proveedores; 5) solicitud de ofertas; 6) selección del proveedor; 7) pedido; 8) evaluación del servicio. Decide un **centro de compras**: quien **inicia**, quien **usa**, quien **influye**, quien **decide**, quien **compra** o paga y quien **filtra** la información. La demanda es **derivada**: Albada pide comidas porque tiene reuniones.

Tabla 4 · Hogar y oficina ante el mismo encargo

Rasgo	Hogar del barrio	Oficina (Albada Asesores)
Necesidad principal	Cenar rápido entre semana	Comer o reunirse sin salir
Quién decide	Una o dos personas	Centro de compras
Frecuencia	Variable, a veces impulsiva	Planificada, a veces fija
Qué exige	Precio, horario, sabor	Factura, puntualidad, alérgenos
Riesgo percibido	Que no guste	Quedar mal en una reunión

Ejemplo resuelto. La compra de Albada Asesores. Mencía, responsable de servicios generales, **inicia**: los jueves hay reunión a mediodía y el equipo sale a comer fuera. **Especifica** (18 raciones, factura, opciones para alergias), pide **ofertas** a La Despensa y a Cocinas Brañas y **filtra** lo que llega a la gerencia, que **decide**; administración **paga** y las 18 personas **usan**. Si la primera entrega llega tarde, la **evaluación** hunde la recompra aunque el kit esté bueno.



ACTIVIDADES

9. Compara las necesidades de un hogar con poco tiempo y de una oficina como Albada. Clasifícalas con la jerarquía de Maslow y ordénalas por prioridad para la decisión de compra, justificando el orden con un hecho del caso.

10. Representa el proceso de compra de un kit por un hogar y por Mecanizados Arenal (taller de unas 60 personas a turnos, sin comedor): fases, papeles del centro de compras y una variable interna o externa en cada caso.

11. En la compra de Albada, ¿qué pregunta harías a Mencía, cuál a la gerencia y cuál a quienes comen el kit? Explica por qué no sirve la misma para todos.

1.5 Segmentación y estrategia

Segmentar es construir grupos con reglas claras; describirlos es medirlos; elegir a cuál dirigirse es otra decisión.

Finalidad, requisitos y criterios. Segmentar es dividir un mercado en grupos homogéneos por dentro y distintos entre sí, para **adaptar la oferta** a cada uno y **usar mejor los recursos**. Un segmento útil es **medible**, **accesible**, **sustancial** (bastante grande para que compense), **diferenciable** y **accionable**. Los criterios pueden ser geográficos (distancia), demográficos, conductuales (frecuencia de pedido) o de necesidad (rapidez, salud, grupo). Las reglas tienen que ser **operativas** (con la misma tabla, cualquiera obtiene los mismos grupos) y **sin solapamientos** (cada registro, en un solo grupo). Fija antes el orden: si una oficina con 0 pedidos se mirara primero como oficina, caería en O y no en N.



Tabla 5 · Registros ficticios de clientes y contactos de kits

(muestra de trabajo, enero de 2027)

ID	Tipo · distancia (m)	Pedidos al mes	Necesidad · franja
C01	hogar · 250	4	rapidez · 18-20 h
C02	hogar · 400	2	saludable · 18-20 h
C03	oficina · 600	6	grupo · 13-14 h
C04	hogar · 1.200	1	saludable · 20-21 h
C05	oficina · 300	8	grupo · 13-14 h
C06	hogar · 350	3	rapidez · 20-21 h
C07	hogar · 1.500	0	rapidez · —
C08	oficina · 900	4	saludable · 13-14 h
C09	hogar · 200	5	rapidez · 18-20 h
C10	hogar · 800	1	saludable · 18-20 h

ID	Tipo · distancia (m)	Pedidos al mes	Necesidad · franja
C11	oficina · 1.300	2	grupo · 13-14 h
C12	hogar · 450	2	rapidez · 20-21 h

Reglas, en este orden. 1) 0 pedidos en el mes → **N** (sin compra en el periodo). 2) Oficina → **O**. 3) Hogar a 500 m o menos con «rapidez» → **H1** (cercanos con prisa). 4) Resto de hogares → **H2** (saludable o lejos).

Tabla 6 · Resultado de la segmentación

Segmento	Registros	Pedidos al mes	Pedidos por registro
O · oficinas	4	20	5,0
H1 · cercanos con prisa	4	14	3,5
H2 · saludable o lejos	3	4	1,3
N · sin compra	1	0	0,0
Total	**12**	**38**	**3,2**

Las oficinas son 4 de 12 registros (33,3 %), pero hacen el 52,6 % de los pedidos (20 de 38); H1, el 36,8 %, y H2, el 10,5 %. Son 12 registros de trabajo, no una muestra del barrio; y la franja «13-14 h» es la hora de comer declarada, mientras que la entrega a oficinas es de 11:30 a 13:30: un dato que hay que revisar.

Describir antes de elegir. Cada segmento se describe con las mismas variables: cuántos registros tiene, cuántos pedidos hace, en qué franja pide y qué necesidad declara. Sin esa ficha, la elección se hace por intuición.

Elegir estrategia. Indiferenciada: una oferta para todos. **Diferenciada:** una para cada segmento elegido (K04 de cena para H1 con reparto de tarde y K03 para O a mediodía). **Concentrada:** todo en un segmento (solo oficinas). La **personalizada**, que ya conoces, queda para encargos a medida. Deciden los **recursos** (la zona de kits prepara oficinas de 7:00 a 11:00 y hogares de 15:30 a 18:00) y las **necesidades**, no el tamaño del segmento.

CLAVE · SEGMENTAR NO ES ELEGIR

Que las oficinas hagan más pedidos en esta tabla no obliga a una estrategia concentrada: antes hay que saber si el segmento es sustancial en el mercado real, si la zona de kits puede servirlo y qué riesgo tiene depender de pocos clientes.

ACTIVIDADES

12. Aplica las reglas a la tabla 5 y comprueba la tabla 6. Cambia la regla 3 a «hogar a 400 m o menos con rapidez» y di qué registro cambia de grupo.
13. Compara para el caso las estrategias indiferenciada, diferenciada y concentrada: oferta, recursos de la zona de kits y riesgo. Elige una y justifícala.
14. Explica por qué la tabla 6 no permite afirmar que «el 52,6 % de la demanda de kits del barrio es de oficinas». ¿Qué tendrías que medir para poder decirlo? Comprueba además que los segmentos O y H1 cumplen los cinco requisitos.

CASO RESUELTO · PASO A PASO

Kits para familias u oficinas

Caso ficticio. Del lunes 11 al viernes 22 de enero de 2027, el equipo de análisis ordena lo que Nora sabe y lo que supone, y le entrega una pregunta de decisión medible.

Paso 1. La decisión, los datos y el alcance. Decisión: a quién dirigir los kits (hogares, oficinas o ambos) y con qué servicio de pedido, antes del 1 de marzo. Quién decide: Nora. Datos: comentarios de los habituales (Olvido, sin registro), precio del K03 (20,00 € sin IVA), horarios de preparación y entrega, supermercado a 300 m y Cocinas Brañas en el polígono. Alcance: barrio y polígono, días laborables de 2027. Falta: datos de quienes no son clientes.

Paso 2. Clasificar los cuatro factores

Factor	Nivel	Control	Qué se hace con él
Precio propio del kit	Empresa (marketing mix)	Controlable	Decidirlo con datos
Competidor nuevo	Microentorno	No controlable	Vigilarlo y diferenciarse
Renta del barrio	Macroentorno (económico)	No controlable	Medirla (INE) y adaptarse
Horarios de los clientes	Conducta del consumidor	No controlable	Medirlos en el estudio

El criterio es la **capacidad de actuación**: solo el precio cambia con una decisión de Nora; el competidor influye, pero no se controla; la renta se toma de una fuente oficial y los horarios hay que averiguarlos.

Paso 3. Comprobar e interpretar. Lo que el análisis **permite** concluir: que la decisión depende de dos datos que La Despensa no tiene (cuántos hogares y oficinas comprarían y en qué franja) y que la opinión de los habituales no los responde. Lo que **no** permite concluir: ni que los kits gustarán a todo el barrio ni que las oficinas sean el mejor segmento. Por eso la salida no es una respuesta, sino una **pregunta medible**, con población, variable y periodo, que no da por buena la opinión de partida.

RESULTADO

Precio propio: variable controlable. Competidor nuevo: microentorno. Renta del barrio: macroentorno. Horarios de los clientes: conducta que hay que medir. Pregunta de decisión entregada a Nora el viernes 22/01/2027: «¿Qué proporción de los hogares del barrio y de las oficinas del barrio y del polígono compraría un kit al menos una vez al mes, a qué precio y con qué forma de entrega?». Es una pregunta medible, no la confirmación de lo que ya se pensaba.

Comprueba que lo has entendido. a) ¿Por qué la renta del barrio no es controlable aunque La Despensa pueda adaptar sus precios a ella? b) ¿Qué parte de la pregunta final no podría responder una encuesta solo a clientes con tarjeta? c) ¿Por qué los horarios no son una variable del marketing mix?

PRÁCTICA

Ruedalia estudia el servicio a empresas

Caso ficticio. Txema, dueño de Ruedalia, quiere ofrecer mantenimiento de bicicletas de empresa en las instalaciones de los clientes del polígono. Antes de dar precio a nadie, te pide que ordenes los factores con el mismo procedimiento que usaste en La Despensa. Datos de partida (febrero de 2027):

- **Capacidad del taller.** Ruedalia abre de lunes a viernes, de 9:00 a 17:00 (40 h a la semana). Ousmane es el único mecánico; según la agenda de las cuatro semanas del 1 al 26 de febrero, tiene ocupadas de media 34 h a la semana en reparaciones del taller. La persona a media jornada hace recogidas y recepción, no repara.
- **Primer cliente.** Mecanizados Arenal, S.L. tiene 14 bicicletas de empresa y pide una revisión básica mensual en su nave. Ousmane calcula 40 minutos por bicicleta, sin contar el desplazamiento.
- **Regulación.** El ayuntamiento tramita una ordenanza que pediría a las empresas del polígono aparcamientos seguros para bicicletas (supuesto ficticio para esta práctica).
- **Compra.** En Arenal propone el jefe de mantenimiento, aprueba la gerencia, paga administración y usan las bicicletas unas 14 personas de los turnos.

Encargo. 1) Clasifica la capacidad del taller, la ordenanza y la compra de Arenal (nivel, control y qué puede hacer Ruedalia). 2) Calcula las horas al mes que exige Arenal y compáralas con la capacidad libre de Ousmane (considera 4 semanas al mes y la revisión repartible), con unidades. 3) Describe el proceso de compra de Arenal con sus decisores. 4) Formula una pregunta de decisión medible para Txema.

Plantilla IC-P01 · Servicio a empresas de Ruedalia

Campo	Tu respuesta
Datos y fuentes con unidad y periodo	
Procedimiento o cálculo	
Resultado y evidencia	
Decisión y limitaciones	

Lista de cotejo

☐ Cada factor con nivel y control.
☐ Horas con unidad y periodo (semana o mes).
☐ Supuestos del cálculo escritos.
☐ Cuatro papeles de la compra de Arenal.
☐ Pregunta medible, no una conclusión.

LECTURA Y DEBATE

Lo que dicen los de siempre



Relato ficticio. El lunes 11 de enero, Olvido salió del obrador con las manos todavía blancas de harina. «Los habituales me dicen que les gusta la idea de los kits. La señora del cuarto me ha pedido uno para el sábado.» Nora tomó la primera decisión allí mismo: el sábado 16 prepararían 30 kits K02 Merienda en familia y los pondrían junto al mostrador del pan. Si se vendían, era la prueba.

El sábado se vendieron 26 de los 30. Olvido estaba contenta. El lunes 18, Adrián repasó los tiques del sábado, uno por uno, y anotó en una hoja cuántos kits iban en compras con tarjeta de la tienda: 22 de los 26; los otros 4, sin tarjeta. Ninguna oficina; no se sabe si por no conocer los kits o porque el sábado cierran.

«Entonces, ¿ha funcionado o no?», preguntó Nora. Adrián le enseñó la hoja: «Ha funcionado con los que ya venían». Olvido se quedó pensando. «Mis clientes no mienten. Pero yo solo hablo con los que entran.» Nora guardó la hoja en la carpeta del estudio y escribió en la portada, con su letra grande: «¿Y los demás?».

Preguntas. 1) ¿Qué afirmación de la historia necesita una comprobación antes de convertirse en decisión? 2) ¿Qué información pedirías primero y por qué cambiaría la respuesta? 3) Propón una actuación provisional que permita aprender sin ocultar la incertidumbre.

APRENDE CON LA IA

Unai pregunta a un asistente de IA cómo conocer a los posibles clientes de los kits. Esta es la respuesta:

RESPUESTA DE LA IA

«Entrevistar clientes actuales basta para conocer a todos los clientes potenciales. Como ya compran en la tienda, representan bien al barrio y a las oficinas. Si diez clientes habituales dicen que les gustan los kits, puedes suponer que el mercado está validado.»

Tu tarea. Localiza el error principal y los que se derivan de él, contrasta cada uno con lo que has trabajado en la unidad (delimitación del mercado, micro y macro, segmentación) y escribe una respuesta corregida que diga también su límite. Conserva la prueba de cada corrección: la tabla, la fuente o el dato del caso que la justifica.

Ojo con la IA. Una respuesta segura no es una respuesta comprobada. Pide siempre a quién se refiere un dato y quién queda fuera.

ACTIVIDADES FINALES

Comprueba

15. Resume en tres líneas la diferencia entre influir en un factor y controlarlo, y aplícala a los cuatro factores del caso resuelto.
16. ¿Por qué micro y macro no se distinguen por el tamaño del agente? Pon un ejemplo del caso.
17. Explica la diferencia entre construir segmentos, describirlos y elegir a cuál dirigirse.

Aplica

18. Repite la práctica de Ruedalia (4 semanas al mes) cambiando una condición: además de Arenal, otras dos empresas piden lo mismo con 10 bicicletas cada una. Recalcula las horas al mes y di qué cambia (el margen de capacidad) y qué se mantiene (la clasificación de los factores).
19. Añade al mapa de la tabla 2 un factor tecnológico y un factor medioambiental, cada uno con su efecto sobre la decisión, su indicador y su fuente.
20. Con la tabla 5, propón una regla de segmentación distinta (por ejemplo, por franja horaria) y comprueba que no tiene solapamientos ni registros sin grupo.

Investiga

21. Localiza una fuente primaria o documentación técnica pertinente para segmentar a los clientes de una tienda de proximidad (una estadística oficial, un manual de una asociación profesional o un informe metodológico). Registra la fecha de consulta, el ámbito, la afirmación que sustenta y una limitación para el caso.
22. Busca en el Atlas de distribución de renta de los hogares del INE la renta media por hogar de tu municipio y de dos secciones censales. Anota año de referencia, fecha de publicación y una razón por la que ese dato no basta para decidir un precio.

TU INVESTIGACIÓN AVANZA · PASO 1 · I01 · MERCADO Y PREGUNTA DE DECISIÓN

Elige el negocio que vas a investigar durante el curso (real con permiso o ficticio) y abre tu dosier. En esta entrega incluye: 1) el mercado delimitado (producto o necesidad, clientes, territorio y periodo); 2) el mapa de variables controlables, micro y macro, con efecto, indicador y fuente; 3) los segmentos que propones con sus reglas; 4) la pregunta de decisión medible. Añade la evidencia (fuentes con fecha), la decisión que apoya, quién es responsable y la fecha de revisión. Si cambias algo más adelante, guarda la versión anterior y el motivo.

Comprueba tu investigación: la pregunta tiene población, variable y periodo; ningún factor externo aparece como controlable.

☐ Hecho ☐ Revisado ☐ Guardado

OJO PROFESIONAL

Documento: ficha de entorno de los kits, versión 0.1

Documento ficticio de La Despensa de Nora, S.L., redactado deprisa por Unai el lunes 18/01/2027 para la reunión con Nora. Contiene **cinco errores**.

N.º	Apartado	Texto del borrador
1	Variables controlables	Precio de los kits, horarios de preparación y renta del barrio
2	Microentorno	Clientes, Huertas del Ribazo, Envaserde, supermercado y Cocinas Brañas
3	Macroentorno	La ordenanza o norma de horarios se anota como dato interno de la tienda
4	Fuente de renta	INE, Atlas de distribución de renta de los hogares, sección censal, año 2024
5	Método para conocer a los no clientes	Preguntar a los compradores con tarjeta qué creen que piensan los demás
6	Segmentos	H1 (hogares con prisa) y H2 (hogares cercanos): un hogar cercano con prisa cuenta en los dos
7	Estrategia	Concentrada en oficinas, porque es el segmento que más pedidos hace en la tabla de trabajo
8	Pregunta del estudio	¿Qué proporción de hogares y oficinas del barrio y del polígono compraría un kit al mes, a qué precio y con qué entrega?

Encargo. Señala las cinco filas con error, explica por qué lo son y reescríbelas para que la ficha se pueda usar. Indica qué filas están bien y por qué.

N.º	Línea	Por qué es un error	Corrección
1			
2			
3			
4			
5			

AUTOEVALUACIÓN

1. La Despensa sube el precio del K03 y el supermercado abre platos preparados. ¿Qué es controlable?
a) El precio.
b) El precio y el supermercado, porque influye en sus ventas.
c) Ninguno de los dos.

2. Un mapa del microentorno es completo cuando...
a) nombra clientes, proveedores, competidores e intermediarios.
b) da la relación y un efecto concreto de cada agente sobre la decisión.
c) ordena los agentes por tamaño.

3. Para saber cuántas empresas hay en el polígono, la fuente adecuada es...
a) el barómetro del CIS.
b) la AEPD, porque registra a las empresas que tratan datos.
c) el censo público de empresas de la cámara de comercio.

4. Una oficina y un hogar piden el K03. ¿Qué es correcto?
a) Ambos siguen la misma jerarquía de necesidades.
b) En la oficina hay que conocer a quien propone, aprueba, paga y usa.
c) En el hogar no hay influencias.

5. En la tabla 5, C06 (350 m, «rapidez», 3 pedidos) va, con las reglas de 1.5, a...
a) H1.
b) H2.
c) H1 y H2.

6. En el caso resuelto, los horarios de los clientes son...
a) una variable del marketing mix.
b) macroentorno legal.
c) conducta del consumidor que hay que medir.

7. En la práctica de Ruedalia, una ordenanza municipal en tramitación es...
a) un dato interno.
b) un factor del macroentorno, no controlable.
c) microentorno, porque es del mismo municipio.

8. Según la IA, los clientes actuales bastan para conocer a los potenciales. Falla porque...
a) deja fuera a quienes no compran y no define la población.
b) diez son pocos; cien bastarían.
c) solo valen fuentes oficiales.

9. En el Ojo profesional, la renta del barrio figura como controlable. Se corrige...
a) pasándola al macroentorno económico, no controlable.
b) pasándola al microentorno.
c) quitándola.

10. Tu I01 dice: «Pregunta: demostrar que el negocio gustará». ¿Qué te obliga a revisarla?
a) Nada: es clara.
b) Tener pocos seguidores en redes.
c) Que no tiene población, variable ni periodo y da por hecha la respuesta.

Ya sé...

- ☐ explicar las distinciones de los cinco apartados: control, entorno, fuentes, compra y segmentación.
- ☐ ejecutar el procedimiento del caso: clasificar factores y formular una pregunta medible.
- ☐ transferirlo a otro negocio, como el servicio a empresas de Ruedalia.
- ☐ detectar una conclusión no sustentada sobre clientes que no se han estudiado.
- ☐ entregar una evidencia revisable, con fuente, fecha y versión.



UNIDAD 2

Sistema de información de marketing

Bloque 1 · Mercado y diseño de la investigación

EN ESTA UNIDAD APRENDERÁS A…

- Partir de las decisiones y sus plazos para determinar qué información necesita una empresa y cuál sobra.
- Clasificar datos por su naturaleza, su origen, su disponibilidad y su fuente, sin confundir las dimensiones.
- Diseñar un sistema de información de marketing con sus subsistemas, sus responsables y su flujo hasta la decisión.
- Juzgar la calidad de un dato (unidad, periodo, cobertura, actualidad) y ajustar el sistema al tamaño de la empresa.
- Convertir registros en indicadores y números índices con fórmulas visibles, registro persistente y consultas guardadas.

SITUACIÓN DE PARTIDA

Lunes 25 de enero de 2027. La semana pasada, del lunes 18 al sábado 23, La Despensa probó el K02 Merienda en familia (11,00 € sin IVA). Nora pide «las ventas de la semana» y recibe tres respuestas distintas: la caja dice **1.306,80 €**; Unai, desde el panel de la web, dice **63 pedidos**; Itxaso, con la hoja del obrador, dice **108 kits**.

Nora, impaciente: «¿Cuál de las tres es la buena?». Antes de decidir si repite la prueba con más kits, te pide averiguar qué mide cada cifra y diseñar un sistema para que desde el lunes 8 de febrero no vuelva a pasar. Lo quiere el viernes 5 de febrero.

PARA EMPEZAR

1. ¿Qué tiene que decidir Nora con «las ventas de la semana»? ¿Necesita euros, pedidos o kits?
2. ¿Qué sabemos de cada cifra (periodo, unidad, fuente) y qué estamos suponiendo al compararlas?
3. ¿Qué comprobación demostraría si las tres cifras son compatibles o si alguna está mal?



Necesidades y decisiones

2.1



Fuentes y datos

2.2



Subsistemas del SIM

2.3



Calidad y periodicidad

2.4



Registro y consulta

2.5

2.1 Necesidades y decisiones

Un dato sin decisión es un coste. Se empieza por lo que hay que decidir, cuándo y quién, y solo después se pregunta qué hay que saber.

De la decisión al dato. Ya sabes que la información sirve para decidir y que saber cuesta. En grado superior el procedimiento se escribe: para cada **decisión** se anota la **pregunta** que hay que responder, el **dato** que la responde (con su unidad), la **fuente** y el **plazo** en que hace falta. Una **necesidad de información** es precisamente eso: la distancia entre lo que se sabe y lo que hay que saber para decidir a tiempo.

Tabla 1 · Decisiones de los kits y sus necesidades de información

Decisión	Pregunta	Dato y unidad	Plazo
Surtido: qué kits mantener	¿Qué kits se repiten?	Kits por código y semana (unidades)	Cada lunes
Precio del K02	¿Se vende a 11,00 €?	Kits vendidos y precio del competidor (€)	Antes de cada prueba
Horario de recogida	¿A qué hora recogen?	Pedidos por franja (pedidos)	Antes de marzo

Ni más ni menos. Recoger «todo por si acaso» aumenta el coste (tiempo de registro, almacenamiento, protección de datos personales) y no garantiza comprender nada: cuantos más datos sin uso, más difícil encontrar el que importa. Lo contrario también es un error: decidir el horario de recogida sin saber a qué hora llegan los pedidos. La pregunta que filtra es siempre la misma: **¿qué decisión cambiaría con este dato?** Si ninguna, no se recoge. Y si la respuesta llega después del plazo de la decisión, tampoco sirve: el dato del horario hace falta antes de marzo, no en abril.

Quién usa la información. Nora decide precio y surtido; Itxaso necesita los kits por día para preparar; Unai, los pedidos por franja para publicar ofertas. Un registro bien definido sirve a los tres, pero cada uno lo consulta con una pregunta distinta: por eso la tabla 1 anota la decisión antes que el dato, y el usuario antes que la herramienta con la que se recoge.

Necesidades periódicas y puntuales. Unas necesidades se repiten con un calendario fijo (los kits de cada lunes, los precios del competidor cada mes) y piden un **registro continuo**; otras aparecen una sola vez (¿abrimos los sábados por la tarde?) y piden un **estudio puntual**. Confundirlas sale caro: un registro diario para una pregunta de una vez, o un estudio semanal para lo que bastaba con anotar.

CLAVE · DATO, INFORMACIÓN Y DECISIÓN

Un **dato** es un registro aislado («63»). La **información** es el dato con su unidad, periodo y contexto («63 pedidos web del 18 al 23/01»). Solo con información se decide. Las tres cifras de Nora eran datos sin etiqueta.

Ejemplo resuelto. Lo que Unai quería guardar. Unai propone registrar en la web, para cada pedido, fecha de nacimiento, dirección completa, red social preferida y número de mascotas. Aplicamos el filtro: la dirección, solo si hay reparto; la franja de entrega, para el horario; la fecha de nacimiento, la red social y las mascotas no cambian ninguna decisión de la tabla 1. Se quedan fuera: no se guardan datos personales sin finalidad.

ACTIVIDADES

1. Asocia cada decisión de surtido, precio y horario de los kits a una pregunta, un dato con su unidad, una fuente y un plazo. Añade una cuarta decisión de tu elección.
2. Itxaso quiere registrar cada día la temperatura de la cocina, los kits preparados por turno y el nombre de quien los prepara. Para cada dato, di qué decisión apoya o explica por qué no lo recogerías.

2.2 Fuentes y tipos de datos

Un mismo dato puede ser interno, cuantitativo y primario a la vez. Las clasificaciones son dimensiones distintas, no casillas que se excluyen.

Tres dimensiones separadas. Ya conoces los tres criterios; aquí se aplican a la vez a cada dato:

- **Naturaleza: cuantitativo** (se cuenta o se mide: kits, euros, minutos) o **cualitativo** (describe opiniones, motivos o conductas: lo que dice una entrevista).
- **Origen o procedencia: interno** (lo genera la empresa: caja, web, obrador) o **externo** (viene de fuera: padrón, informe sectorial, precios del competidor).
- **Disponibilidad: secundario** (ya existe, recogido para otro fin) o **primario** (se recoge expresamente para esta investigación).

Tabla 2 · Cuatro datos clasificados en las tres dimensiones

Dato	Naturaleza	Origen	Disponibilidad
Ventas de la caja	Cuantitativo	Interno	Secundario
Entrevista a Mencía (Albada)	Cualitativo	Externo	Primario
Padrón municipal	Cuantitativo	Externo	Secundario
Observación de la cola del súper	Cuantitativo o cualitativo	Externo	Primario

Fíjate en las ventas de la caja: son **internas** y, aun así, **secundarias** para el estudio, porque se registraron para cobrar, no para investigar. Y la observación de la cola puede dar un recuento (cuantitativo) y unas notas de conducta (cualitativo) en la misma hoja.

Clasificar decide el método. Lo **secundario** es barato y rápido, pero se recogió con otra definición y otra fecha; lo **primario** se ajusta a la pregunta, pero cuesta tiempo y dinero. Por eso se busca primero lo secundario y solo se recoge lo primario que falta.

La fuente, por su procedencia. Además, cada dato tiene una **fuente**: la persona, el sistema o el organismo del que procede. Distingue la **fuente original**, que produce el dato (el ayuntamiento con el padrón; el TPV con la venta), de la **fuente intermediaria**, que lo repite o lo resume (una noticia sobre el padrón; lo que Unai copia del TPV). Se cita la original con su fecha y sus **condiciones de acceso** (pública, con usuario, interna, con datos personales).

CLAVE · DIMENSIONES, NO CASILLAS

«¿Es primario o interno?» está mal preguntado: puede ser las dos cosas. Clasifica en las tres dimensiones y anota la fuente original.

Tabla 3 · Diez registros que llegan a la mesa de Nora

(enero de 2027)

N.º	Registro	Quién lo aporta
1	Cierre de caja del sábado 23/01	TPV (Nora)
2	Pedidos web de la semana	Panel de la web (Unai)
3	Hoja de kits preparados	Obrador (Itxaso)
4	Libreta de peticiones del mostrador	Olvido
5	Precios del supermercado del lunes 18/01	Visita de Adrián
6	Población del municipio	INE, en una noticia del periódico local
7	Reseñas de la tienda en internet	Unai
8	Notas de la entrevista con Mencía	Equipo de análisis
9	Altas de la tarjeta de cliente	TPV (Nora)
10	Informe de una asociación de comerciantes	Correo de la asociación

ACTIVIDADES

3. Clasifica las ventas de la caja, la entrevista a Mencía, el padrón y la observación de la cola del supermercado por naturaleza, origen y disponibilidad, y explica un caso en que el mismo dato cambie de disponibilidad según el estudio.
4. Etiqueta los registros de la tabla 3 con las tres dimensiones, fuente original o intermediaria, fecha y acceso. ¿Cuál sustituirías por su fuente original?

2.3 Subsistemas del SIM

Un sistema de información no es un programa ni una carpeta: son personas, reglas y registros que llevan un dato hasta una decisión.

Qué es un SIM. Un **sistema de información de marketing (SIM)** es el conjunto de personas, procedimientos y herramientas que **recogen, ordenan, analizan y distribuyen** información oportuna y exacta para las decisiones de marketing. Ya conoces sus cuatro subsistemas; aquí se les asignan funciones, entradas y responsables:

Tabla 4 · Subsistemas del SIM de La Despensa

Subsistema	Función	Entradas	Responsable
Datos internos o registros internos	Registrar lo que pasa en la empresa	Caja, web, obrador	Itxaso
Inteligencia de marketing	Vigilar el entorno de forma continua	Precios, aperturas, normas	Unai
Investigación comercial	Responder una pregunta concreta	Entrevistas, encuestas	Equipo de análisis
Análisis o apoyo a las decisiones	Convertir datos en indicadores	Hojas, consultas	Equipo de análisis

La investigación comercial dentro del SIM. La **investigación comercial** es la función que diseña, recoge, analiza y comunica de forma sistemática la información necesaria para una **decisión comercial concreta**. Su **finalidad** es reducir la incertidumbre de esa decisión con datos fiables, no confirmar lo que ya se piensa. Se diferencia de los otros subsistemas en que es **puntual** (empieza con una pregunta y acaba con un informe) y **propia** (recoge datos primarios si los existentes no bastan). Sus **aplicaciones** cubren todas las decisiones de marketing: estudios de **mercado** (tamaño, segmentos), de **consumidor** (motivos, hábitos), de **producto** (pruebas, envase), de **precio** (sensibilidad), de **distribución** (canales, horarios), de **comunicación** (recuerdo de un anuncio) y de **competencia**. El estudio de kits de 2027 es investigación comercial; la visita mensual al supermercado, inteligencia.

Inteligencia sin espionaje. La inteligencia de marketing se alimenta de **fuentes lícitas**: los precios a la vista en el lineal, la web del competidor, la prensa local, el BOE y el boletín de la comunidad autónoma, las ferias y lo que comentan los clientes en el mostrador. Lo útil es la **continuidad**: una ficha con la misma estructura cada mes permite ver cambios; una visita suelta, no. La apertura de una cafetería con obrador de cadena el 01/02/2027 es una noticia de inteligencia que Unai debe registrar con su fecha y su fuente.

El flujo del dato a la decisión. El recurso técnico de esta unidad es el flujo **dato → procesamiento → información → decisión**, con su responsable en cada paso:

Tabla 5 · Flujo de las ventas semanales de kits

Paso	Qué ocurre	Quién	Cuándo
Dato	Venta en caja, pedido web, kit preparado	Caja, web, obrador	Cada día
Procesamiento	Se juntan las tres fuentes por día y canal	Itxaso	Lunes, 9:00
Información	Kits, pedidos e importe sin IVA, con definición	Análisis	Lunes, 11:00
Decisión	Kits que se preparan la semana siguiente	Nora	Lunes, 13:00

Qué datos procesa. Un SIM procesa **variables** definidas, cada una con su **unidad**, su **periodicidad** y sus **relaciones** con otras: el pedido web (pedidos) se relaciona con los kits (unidades) por el número de pedido; el kit, con el cliente por la tarjeta, si la tiene; el precio del competidor (€), con el kit equivalente.

EN TU SECTOR

Moda: la inteligencia vigila tendencias y escaparates de la competencia; la investigación prueba una colección con clientes. **Turismo:** los registros internos son reservas y ocupación; la investigación, una encuesta de satisfacción tras la estancia. **Banca:** el apoyo a las decisiones calcula qué clientes pueden necesitar un producto, con los límites de la protección de datos.

ACTIVIDADES

5. Diseña el flujo de registros internos, inteligencia, investigación y apoyo analítico hacia Nora para la decisión de horario de recogida: qué entra en cada subsistema, quién lo procesa, qué información sale y cuándo llega a la decisión.
6. Define las variables de ventas, clientes y competencia del SIM: nombre, unidad, periodicidad, fuente y relación con otra variable (al menos dos por área).
7. Clasifica como investigación comercial, inteligencia o registro interno: a) el cierre de caja diario; b) una encuesta a oficinas sobre su hora de comer; c) la visita mensual al supermercado; d) una prueba del K04 con diez familias. Para las investigaciones, indica su tipo de aplicación.

2.4 Calidad y periodicidad

Un dato es útil si se sabe qué mide, de cuándo es, a quién cubre y quién responde de él. La sofisticación del sistema se ajusta al tamaño de la tienda y al valor de la decisión.

Cinco atributos de calidad. Antes de publicar un indicador se comprueban: **definición y unidad** (¿kits, pedidos o euros?, ¿con o sin IVA?); **periodicidad** (cada cuánto se actualiza y qué periodo cubre: «semana de lunes a sábado»); **cobertura** (qué canales, qué clientes: ¿la caja incluye los pedidos web?); **actualidad** (fecha de la última carga); y **responsable** (quién lo mantiene y responde de él). Un dato exacto pero de otra semana, o con IVA cuando el resto va sin IVA, induce a error. Estos atributos se escriben en el **diccionario de datos**: una tabla con cada variable, su definición, su unidad, su fuente y su responsable, que cualquier persona del equipo puede consultar antes de usar una cifra. Sin diccionario, cada uno interpreta «ventas» a su manera, como pasó con Nora.

LA NORMA DICE

Reglamento (UE) 2016/679 (RGPD), de 27 de abril de 2016, art. 5.1: los datos personales serán «c) adecuados, pertinentes y limitados a lo necesario en relación con los fines para los que son tratados ("minimización de datos")» y «d) exactos y, si fuera necesario, actualizados». Consulta la versión vigente.



Ejemplo resuelto. Las tres cifras conciliadas. Adrián compara las tres fuentes línea a línea y el equipo comprueba las cuentas:

Tabla 6 · Conciliación de la semana del 18 al 23/01/2027

Fuente	Cifra	Unidad y cobertura	Comprobación
A · Caja (TPV)	1.306,80 €	Euros con IVA, todos los canales	1.306,80 ÷ 1,10 = 1.188,00 € sin IVA
B · Panel web	63	Pedidos, solo web	84 kits ÷ 63 pedidos = 1,33 kits por pedido
C · Obrador	108	Kits, todos los canales	84 web + 24 tienda = 108

Las tres son ciertas y compatibles si los 108 kits preparados se vendieron todos, sin mermas: 108 kits × 11,00 € = 1.188,00 € sin IVA, que con el 10 % de IVA son los 1.306,80 € cobrados. Lo que no se puede hacer es sumarlas ni compararlas sin unidad: «63 frente a 108» no es una diferencia de ventas, son cosas distintas.

Sencillo o complejo. Un SIM se ajusta al **tamaño** de la empresa, al **volumen** de datos, a la **frecuencia** de las decisiones y al **valor** de lo que se decide. Para una tienda de barrio, una hoja de cálculo compartida, el TPV y un procedimiento escrito pueden bastar; un programa de gestión de clientes con cuadros de mando automáticos tiene sentido si hay muchos pedidos diarios y decisiones frecuentes. Más complejidad no es más calidad: cuesta dinero, formación y mantenimiento, y si nadie la mantiene acaba abandonada.

Qué periodicidad. La periodicidad del indicador sigue a la de la decisión: los kits se deciden cada lunes, así que el indicador es **semanal**; el precio del competidor se revisa cada mes, así que la observación es **mensual**. Actualizar más a menudo de lo que se decide solo añade trabajo; actualizar menos obliga a decidir con datos viejos. Además, el periodo tiene que ser siempre el mismo: si una semana va de lunes a sábado y otra de martes a lunes, las cifras dejan de ser comparables. La fecha de corte (el sábado a las 14:30, al cerrar la tienda) se escribe también en el diccionario.

ACTIVIDADES

8. Compara para La Despensa un SIM sencillo (TPV, hoja de cálculo compartida y procedimiento escrito) y uno complejo (programa de gestión de clientes y cuadro de mando automático): coste, volumen que soporta, frecuencia de actualización y utilidad para las decisiones de la tabla 1. Selecciona los recursos que propones y justifícalo.
9. Una semana la caja marca 1.452,00 € con IVA y la hoja del obrador 120 kits K02. ¿Son compatibles las dos cifras? Comprueba con el precio sin IVA y el tipo de IVA, y explica qué revisarías si no lo fueran.

2.5 Registro y consulta

El dato tiene que quedar guardado de forma que se pueda actualizar sin duplicarlo y consultar sin rehacer las cuentas.

Archivo, base de datos y presentación. Un **archivo de trabajo** es la hoja donde se calcula; una **base de datos** es un conjunto organizado de **tablas** formadas por **registros** (filas: una venta, un pedido) y **campos** (columnas: fecha, canal, kits), con una **clave** que identifica cada registro (el número de pedido); una **presentación ejecutiva** es el resumen para decidir. Mezclarlos es un error típico: si Nora corrige una cifra en la presentación, el registro sigue mal.

Tabla 7 · Registro diario de la prueba del K02

(hoja del SIM, semana del 18 al 23/01/2027)

Día	Pedidos web	Kits web	Kits tienda
Lunes 18	9	12	2
Martes 19	10	13	3
Miércoles 20	11	15	3
Jueves 21	12	16	4
Viernes 22	13	17	5
Sábado 23	8	11	7
Total	**63**	**84**	**24**

Fórmulas visibles y consulta guardada. Cada total es una fórmula a la vista: =SUMA(C2:C7) da los 84 kits web; =C8/B8, los 1,33 kits por pedido. Cada semana nueva **añade filas** con su fecha, sin sobrescribir las anteriores. Una **consulta guardada** (filtro, tabla dinámica o SQL guardados) repite el mismo resumen cada lunes sin rehacerlo. En SQL estándar, con una fila por día y canal: `SELECT canal, SUM(kits) FROM ventas WHERE fecha BETWEEN '2027-01-18' AND '2027-01-23' GROUP BY canal;`. En la unidad 11 la base será relacional.

Números índices. Para seguir una variable en el tiempo se usa un **número índice**: compara el valor de un periodo con el de un **periodo base**, al que se da el valor 100. El **índice simple** es I = (valor del periodo ÷ valor del periodo base) × 100. Con la semana del 18 al 23/01 como base (108 kits), las tres semanas siguientes, registradas hasta el 13/02, dieron 99, 117 y 126 kits:

Tabla 8 · Índice de kits K02 preparados

(base: semana del 18/01 = 100)

Semana	Kits	Índice	Tasa sobre la semana anterior
18-23/01	108	100,0	—
25-30/01	99	91,7	−8,3 %
01-06/02	117	108,3	+18,2 %
08-13/02	126	116,7	+7,7 %

La **tasa de variación** entre dos periodos es (valor final − valor inicial) ÷ valor inicial × 100: de 99 a 126 kits (semanas del 25/01 y del 08/02), (126 − 99) ÷ 99 × 100 = 27,3 %. Un **índice de precios** aplica la misma fórmula a un precio: si el saco de 25 kg de harina de Molinos Peñarrubia costaba 14,50 € en enero de 2026 y 15,95 € en enero de 2027, el índice es 15,95 ÷ 14,50 × 100 = 110,0 (base enero de 2026 = 100): un 10,0 % más caro. El IPC del INE combina muchos precios ponderados.

CLAVE · PUNTOS NO ES LO MISMO QUE PORCENTAJE

Del índice 91,7 al 116,7 hay 25,0 puntos, pero la variación es del 27,3 %, porque se calcula sobre el valor de partida. Di siempre la base del índice y si hablas de puntos o de porcentaje.



ACTIVIDADES

10. Con la tabla 7, prepara el resumen semanal por canal: kits, pedidos web e importe sin IVA de cada canal (todos son K02 a 11,00 €), y el total conciliado con la caja. Interpreta en dos líneas para la decisión de Itxaso.

11. Crea en una hoja de cálculo (Sheets, Calc u otra) la tabla inicial del SIM con la tabla 7 reorganizada con una fila por día y canal, añade la semana del 25 al 30/01 con 99 kits (inventa y anota su reparto) y guarda una consulta que dé los kits por semana y canal. Las fórmulas, visibles.

12. Calcula el índice de la semana del 08/02 con base en la semana del 25/01 y la tasa de variación entre la primera y la última semana de la tabla 8. Explica por qué cambia el índice y no la tasa.

CASO RESUELTO · PASO A PASO

El SIM de los kits

Caso ficticio. Del lunes 25 de enero al viernes 5 de febrero de 2027, el equipo de análisis aclara las tres cifras y diseña el SIM que Nora usará cada lunes.

Paso 1. La decisión, los datos y el alcance. Decisión: cuántos kits preparar cada semana y si se repite la prueba con más kits. Datos: caja (euros con IVA, todos los canales), panel web (pedidos web) y hoja del obrador (kits). Alcance: semanas de lunes a sábado, kits del catálogo, tienda y web; todo sin IVA salvo aviso.

Paso 2. Separar los subsistemas y definir los indicadores

Fuente	Subsistema	Propietario y fecha	Uso
Ventas semanales (caja, web, obrador)	Registros internos	Itxaso, cada lunes	Surtido y preparación
Observación de precios del súper y de Cocinas Brañas	Inteligencia	Unai, primer lunes de mes	Precio
Encuesta a hogares y oficinas	Investigación	Equipo de análisis, una vez	Segmento y horario

Indicadores publicados, cada uno con su definición: **kits vendidos** (unidades, lunes a sábado, todos los canales; fuente: obrador cruzado con caja); **pedidos web** (pedidos, solo web); **importe sin IVA** (euros; caja ÷ 1,10 para kits al 10 %). Comprobación de la semana del 18/01: 108 × 11,00 = 1.188,00 € sin IVA = 1.306,80 ÷ 1,10.

Paso 3. Comprobar e interpretar. El SIM **permite** que el lunes a las 13:00 Nora tenga tres indicadores definidos, conciliados entre sí y con dueño, y una serie semanal que no mezcla unidades. **No permite** saber por qué se vende más o menos ni qué piensan quienes no compran: eso es tarea de la investigación (la encuesta). Tampoco convierte una semana buena en tendencia.

RESULTADO

SIM con tres fuentes separadas: registros internos (ventas semanales, Itxaso, cada lunes), inteligencia (precios, Unai, primer lunes de mes) e investigación (encuesta puntual). El equipo de análisis (apoyo a las decisiones) publica tres indicadores con definición, unidad y fecha de actualización. Las tres cifras de Nora eran ciertas: 1.306,80 € con IVA (1.188,00 € sin IVA), 63 pedidos web (84 kits) y 108 kits en total.

Comprueba que lo has entendido. a) ¿Por qué no se puede decir que «la web vendió 63 y el obrador 108»? b) ¿Qué subsistema responde a la pregunta «a qué hora comen las oficinas»? c) ¿Qué pasaría con el importe si un kit con IVA al 4 % entrara en la caja?

PRÁCTICA

Facturas, agenda y entrevistas de Ruedalia

Caso ficticio. Txema quiere saber qué revisiones repiten sus clientes y qué opinan del servicio. En la semana del lunes 1 al viernes 5 de febrero de 2027 tiene tres registros:

Registro	Cifra	Qué contiene
Agenda de citas del taller	52 citas	Fecha, hora, cliente y tipo de revisión
Facturas emitidas	46 facturas	Cliente, servicio e importe
Entrevistas breves al recoger la bicicleta	9 entrevistas	Opinión sobre plazo y trato

Ousmane explica que 4 citas no se presentaron y 2 fueron reparaciones en garantía, que no se facturan. Esa semana no hubo trabajos sin cita.

Encargo. 1) Clasifica los tres registros (naturaleza, origen, disponibilidad) y separa los de transacción de los de opinión. 2) Propón cómo enlazarlos con un identificador sin duplicar datos de cliente. 3) Comprueba que las cifras cuadran y calcula el porcentaje de citas facturadas y de no presentadas. 4) Propón tres indicadores con definición, responsable y periodicidad, y di qué no permiten concluir las 9 entrevistas.

Plantilla IC-P02 · SIM de Ruedalia

Campo	Tu respuesta
Datos y fuentes con unidad y periodo	
Procedimiento o cálculo	
Resultado y evidencia	
Decisión y limitaciones	

Lista de cotejo

☐ Transacción y opinión separadas.
☐ Identificador común sin duplicar datos.
☐ Cuadre 52 = 46 + 4 + 2.
☐ Porcentajes con su base.
☐ Indicadores con definición, responsable y fecha.

LECTURA Y DEBATE

Raciones o menús



Relato ficticio. Antes de llegar a La Despensa, Itxaso pasó ocho años de jefa de partida en una cocina de catering para colectividades. Un lunes, el encargado de compras abrió el informe de la semana: «640 servicios». Hizo el pedido de pescado para 640 menús del viernes y se fue tranquilo. Era la primera decisión, rápida y con un número en la mano.

El jueves, Itxaso vio entrar el doble de cajas de pescado que de costumbre. Bajó al despacho y pidió el informe. «Servicios», ponía, sin más. Llamó a quien lo había hecho: en ese informe, cada primer plato y cada segundo plato contaban como un servicio. Los 640 servicios eran 320 menús.

Media cámara de pescado se tuvo que congelar o cambiar de plato. Desde entonces, Itxaso no acepta un número sin su unidad. En La Despensa lo repite cada lunes: «Kits, pedidos o euros. Y de qué días». Cuando Nora le enseñó las tres cifras de la semana del K02, sonrió: «Esto ya lo he vivido».

Preguntas. 1) ¿Qué afirmación de la historia necesitaba una comprobación antes de convertirse en decisión? 2) ¿Qué información pedirías primero y por qué cambiaría la respuesta? 3) Propón una actuación provisional que permita aprender sin ocultar la incertidumbre.

APRENDE CON LA IA

Unai pregunta a un asistente de IA cómo montar el sistema de información de La Despensa. Esta es la respuesta:

RESPUESTA DE LA IA

«Un SIM consiste en almacenar todos los archivos en una carpeta compartida. Cuantos más archivos guardes, mejor informado estarás. Si las cifras no coinciden, quédate con la más alta, que suele ser la más completa.»

Tu tarea. Localiza los errores, contrasta cada uno con lo que has trabajado (subsistemas, flujo, calidad, conciliación) y escribe una respuesta corregida que diga también su límite. Conserva la prueba de cada corrección: la tabla o el cálculo de la unidad que la justifica.

Ojo con la IA. Una carpeta guarda archivos; un sistema decide qué se guarda, quién lo mantiene y para qué decisión sirve.

ACTIVIDADES FINALES

Comprueba

13. Resume la diferencia entre registro interno, inteligencia e investigación comercial y aplícala a las tres fuentes del caso resuelto.
14. ¿Por qué las ventas de la caja son internas y, a la vez, secundarias para el estudio de kits?
15. Explica la diferencia entre un índice simple y una tasa de variación con un ejemplo de la tabla 8.

Aplica

16. Repite la práctica de Ruedalia cambiando una condición: la semana siguiente hay 50 citas, 3 no presentadas, 1 en garantía y 2 trabajos sin cita que sí se facturan. ¿Cuántas facturas debería haber? Señala qué resultado cambia y qué procedimiento se mantiene.
17. Ruedalia cobraba la revisión básica a 35,00 € (sin IVA) en enero de 2026 y a 38,50 € en enero de 2027. Calcula el índice de precios con base enero de 2026 y la tasa de variación.
18. Diseña la ficha de un indicador del SIM de La Despensa de tu elección: nombre, definición, unidad, fórmula, fuente, periodicidad, responsable y decisión que apoya.

Investiga

19. Localiza documentación técnica pertinente para registrar y consultar datos (la ayuda oficial de tu hoja de cálculo sobre tablas dinámicas o la de un gestor de bases de datos sobre consultas). Registra la fecha, el ámbito, la afirmación que sustenta y una limitación.
20. Busca en la web del INE el índice general del IPC más reciente: anota su periodo, su base y la tasa de variación anual publicada, con la fecha de consulta.

TU INVESTIGACIÓN AVANZA · PASO 2 · I02 · DISEÑO DEL SIM

Para el negocio de tu investigación, diseña su SIM: 1) las decisiones y sus necesidades de información (decisión, pregunta, dato, fuente, plazo); 2) los cuatro subsistemas con entradas y responsables; 3) el flujo dato → procesamiento → información → decisión; 4) tres indicadores con definición, unidad y periodicidad; 5) una hoja de registro con dos periodos y una consulta guardada. Incluye la evidencia (la hoja con las fórmulas), la decisión que apoya, el responsable y la fecha de revisión. Si cambias algo, conserva la versión anterior y el motivo.

Comprueba tu investigación: ningún indicador mezcla unidades; cada registro tiene fecha y fuente.

☐ Hecho ☐ Revisado ☐ Guardado

OJO PROFESIONAL

Documento: informe semanal de kits, versión 0.1

Documento ficticio de La Despensa de Nora, S.L., preparado por Unai el lunes 25/01/2027 para Nora. Contiene **cinco errores**.

N.º	Apartado	Texto del informe
1	Ventas de la semana	63 pedidos + 108 kits = 171 ventas
2	Importe	1.306,80 € sin IVA
3	Kits por pedido web	84 kits ÷ 63 pedidos = 1,33
4	Hoja del obrador	Sin fecha: «la de esta semana»
5	Precio del K02	11,00 € sin IVA (12,10 € con IVA)
6	SIM	Carpeta compartida «Kits» con todos los archivos
7	Responsable de los datos	Todos
8	Siguiente actualización	Lunes 01/02/2027

Encargo. Señala las cinco filas con error, explica por qué lo son y reescríbelas para que el informe se pueda usar. Indica qué filas están bien.

N.º	Línea	Por qué es un error	Corrección
1			
2			
3			
4			
5			

AUTOEVALUACIÓN

1. Nora decidirá el horario de recogida antes de marzo. ¿Qué dato es necesario?
a) La fecha de nacimiento de los clientes web.
b) Los pedidos por franja horaria.
c) Todos los datos de la web, por si acaso.

2. Las ventas de la caja de La Despensa son, para el estudio de kits...
a) internas y secundarias.
b) internas y primarias.
c) externas y secundarias.

3. La visita mensual de Unai a los precios del supermercado pertenece al subsistema de...
a) investigación comercial.
b) registros internos.
c) inteligencia de marketing.

4. Para una tienda de barrio con unos 100 pedidos de kits a la semana, lo razonable es...
a) un programa de gestión de clientes con cuadro de mando automático.
b) TPV, hoja de cálculo compartida y procedimiento escrito.
c) no registrar nada hasta tener más ventas.

5. Con 99 kits una semana y 126 dos semanas después, la tasa de variación es...
a) 27,3 %.
b) 21,4 %.
c) 27 puntos.

6. En el caso resuelto, 1.306,80 € de caja equivalen a...
a) 1.306,80 € sin IVA.
b) 1.188,00 € sin IVA, 108 kits a 11,00 €.
c) 63 pedidos a 11,00 €.

7. En Ruedalia hubo 52 citas, 46 facturas, 4 no presentados y 2 garantías. Las citas facturadas son...
a) el 88,5 % de las citas.
b) el 46 % de las citas.
c) el 113 % de las facturas.

8. La IA dice que un SIM es una carpeta con todos los archivos. Falla porque...
a) la carpeta debería estar en la nube.
b) un SIM son personas, reglas, flujos y responsables al servicio de decisiones.
c) solo hay que guardar los archivos recientes.

9. En el Ojo profesional, «63 pedidos + 108 kits = 171 ventas» se corrige...
a) separando unidades: 63 pedidos web (84 kits) y 108 kits en total.
b) sumando también los euros.
c) dejando solo la cifra más alta.

10. Tu I02 tiene un indicador «ventas» sin unidad. ¿Qué te obliga a revisarlo?
a) Que el número es pequeño.
b) Que no tiene gráfico.
c) Que no se sabe si cuenta pedidos, unidades o euros, ni de qué periodo.

Ya sé...

- ☐ explicar las distinciones de los cinco apartados: decisión, fuentes, subsistemas, calidad y consulta.
- ☐ ejecutar el procedimiento del caso: conciliar cifras y diseñar un SIM con indicadores definidos.
- ☐ transferirlo a otro negocio, como Ruedalia.
- ☐ detectar una conclusión no sustentada por mezclar unidades o fuentes.
- ☐ entregar una evidencia revisable: hoja con fórmulas, fechas y consulta guardada.



UNIDAD 3

Problema, objetivos y diseño

Bloque 1 · Mercado y diseño de la investigación

EN ESTA UNIDAD APRENDERÁS A…

- Separar síntoma, decisión y problema de información, y reformular preguntas que ya traen la respuesta.
- Formular objetivos medibles, con población, variable y periodo, e hipótesis que los datos puedan desmentir.
- Elegir un diseño exploratorio, descriptivo o causal, y explicar por qué un «antes y después» no prueba un efecto.
- Ordenar las fases de una investigación y elegir fuentes y técnicas por pertinencia, calidad y coste.
- Preparar un plan con fases, responsables, calendario, presupuesto, muestra, análisis y riesgos.

SITUACIÓN DE PARTIDA

Lunes 8 de febrero de 2027. La semana pasada Unai pagó un anuncio en redes de los kits. Nora compara las ventas de kits de esa semana con las de la semana anterior y lo tiene claro: «Han subido. El anuncio ha funcionado». Itxaso no está tan segura: esa semana entró un pedido grande de una oficina e hizo mucho frío.

- Nora quiere lanzar el kit **K03 Comida de trabajo** (20,00 € sin IVA) para las oficinas del barrio y del polígono, y te pide un estudio «que demuestre que funcionará».
- Dispone de **300,00 €** y de **cuatro semanas** (15/02 a 12/03) y decide el **lunes 15 de marzo**. Solo tiene datos internos.

Tu equipo preparará el **plan del estudio** (EK-2027/PL-v1.0) antes de recoger datos nuevos.

PARA EMPEZAR

1. ¿Qué tiene que decidir Nora el 15 de marzo y qué tiene que comprobar antes de hacerlo?
2. ¿Qué sabe La Despensa sobre las oficinas y qué está suponiendo sin haberlo comprobado?
3. ¿Qué dato permitiría revisar la primera respuesta de Nora («el anuncio ha funcionado»)?
4. ¿Qué más pudo cambiar de una semana a otra, además del anuncio?



3.1 Decisión y problema

Una investigación empieza por la decisión que alguien tiene que tomar, no por las ganas de preguntar.

Lo que ya sabes. Las fases y los tres diseños se presentan en 3.3 y 3.4; aquí plantearás bien la primera fase: una pregunta mal planteada produce datos que no sirven.

Síntoma, decisión y problema de información. Conviene separar tres cosas que en la conversación diaria se mezclan:

- El **síntoma** es lo que se observa y preocupa: «vendemos poco a las oficinas», «las ventas subieron después del anuncio». Es un hecho o una impresión, no una pregunta.
- El **problema de decisión** es lo que la dirección tiene que elegir: ofrecer o no el K03 a oficinas, a qué precio, con qué entrega. Lo resuelve quien decide (Nora), no el estudio.
- El **problema de información** (o problema de investigación) es lo que hace falta saber para elegir mejor: qué oficinas comerían un kit preparado, con qué frecuencia, a qué precio y en qué franja. Es lo que resuelve la investigación.

Cuándo hace falta investigar. Hay **necesidad de investigación comercial** cuando hay una decisión pendiente con fecha, existe **incertidumbre** sobre algo que cambiaría la decisión, equivocarse cuesta más que el estudio y la información no está ya disponible. Si Nora fuera a lanzar el K03 pase lo que pase, el estudio sería un gasto.

Qué no resuelven las ventas internas. Los registros de caja y de pedidos describen a quien **ya compra**. No dicen nada de las oficinas que no compran (la mayoría del mercado potencial), no explican por qué se compra o se deja de comprar y no separan el efecto de un anuncio de lo demás que pasó esa semana.

Ejemplo. ¿Cuánto cuesta saber? El K03 (20,00 € sin IVA; coste de referencia 12,00 €) deja un **margen de contribución** de 8,00 € por kit, sin preparación ni reparto. El estudio cuesta 300,00 €: 300,00 ÷ 8,00 = 37,5 → hacen falta **38 kits** como mínimo para recuperarlo. Lanzar sin saber puede costar más: horas, envases y repartos en vano.

Preguntas que ya traen la respuesta. «Demostrar que el cliente quiere nuestro kit» no es una pregunta: es una conclusión que busca pruebas. Un estudio así tiende a elegir a quién preguntar, cómo preguntar y qué datos mirar para confirmar lo que ya se piensa (**sesgo de confirmación**). Una pregunta investigable admite cualquier respuesta, también la negativa, y nombra tres cosas:

- **Población**: de quién se quiere saber algo (oficinas del barrio y del polígono, no «la gente»).
- **Variable**: qué se mide (interés declarado, frecuencia, precio aceptado, franja preferida).

- **Periodo**: cuándo (febrero y marzo de 2027).

Reformulada: «¿Qué proporción de las oficinas del barrio y del polígono declara que encargaría una comida de trabajo preparada al menos una vez al mes, a qué precio y con qué entrega, según una consulta hecha entre el 15/02 y el 12/03/2027?». La palabra **declara** es importante: lo que se mide es una intención, no una compra.



CLAVE · SÍNTOMA, DECISIÓN E INFORMACIÓN

El síntoma se observa, la decisión la toma la dirección y el estudio responde el problema de información. Si la pregunta solo admite un «sí», no es una investigación: es una justificación.

ACTIVIDADES

1. Explica qué incertidumbre no resuelven las ventas internas de kits de La Despensa y qué decisión de Nora cambiaría con la respuesta del estudio del K03.
2. Reformula el síntoma «vendemos poco» como una pregunta investigable para los kits de hogar: escribe la población, la variable y el periodo, y comprueba que admite una respuesta negativa.
3. Clasifica como síntoma, problema de decisión o problema de información: a) «Los sábados se acaban los K02 a las doce»; b) «¿Preparamos más K02 los sábados?»; c) «¿Cuántos hogares reservarían el K02 para el sábado?».

3.2 Objetivos e hipótesis

Un objetivo bien escrito dice qué se va a medir, en quién y cuándo. Una hipótesis dice qué resultado la desmentiría.

Objetivo general y objetivos específicos. El **objetivo general** dice en una frase qué se quiere saber y para qué decisión; los **objetivos específicos** lo dividen en partes medibles, cada una con su fuente, instrumento y análisis. Verbos que el diseño puede cumplir: **estimar, describir, identificar, comparar, explorar**; «demostrar» o «confirmar» prometen el resultado antes de tenerlo, y «saber» o «conocer» no dicen qué se medirá.

Cada objetivo hace visibles la **población** o unidad de análisis (la oficina, no cada empleado), la **variable** con su forma de medirla (interés de 1 a 5, comensales, franja) y el **periodo**. Sin uno de los tres, nadie podrá comprobar si se cumplió.

Ejemplo. Objetivos del estudio del K03

- **Objetivo general.** Estimar el interés declarado de las oficinas del barrio y del polígono por un servicio de comida de trabajo preparada y las condiciones en que lo encargarían, entre el 15/02 y el 12/03/2027, para decidir el 15/03 si se ofrece el K03.
- **OE1.** Identificar los motivos y las barreras de las oficinas para encargar comida preparada (horario, comedor, factura, alergias).
- **OE2.** Estimar la proporción de oficinas que declara que encargaría el K03 al menos una vez al mes a 20,00 € sin IVA.
- **OE3.** Describir la frecuencia prevista, los comensales por pedido, la franja de entrega y la forma de pedido y de pago preferidas.
- **OE4.** Identificar quién propone, quién aprueba y quién paga la compra en cada oficina.

Hipótesis. Una **hipótesis** es una respuesta provisional a una pregunta del estudio, que los datos pueden confirmar o desmentir. Para ser **contrastable** nombra variable, población y el umbral o comparación que la desmentiría:

- H: «Al menos una de cada cuatro oficinas de la zona declara que encargaría el K03 al menos una vez al mes a 20,00 € sin IVA». Se desmiente si la proporción observada, con su margen, queda por debajo del 25 %. Si el 25 % queda dentro del margen, ni se desmiente ni se afirma.

Promesas que el diseño no puede sostener. Con cuatro semanas y 300,00 € no se puede «conocer el mercado de comida de empresa de la ciudad» ni prever las ventas del año; sí estimar el interés declarado de las oficinas cercanas y sus condiciones. Un objetivo prudente que se cumple vale más que uno ambicioso que nadie puede comprobar.

CLAVE · UNA HIPÓTESIS NO ES UN HECHO

Se escribe antes de recoger los datos, se dice qué resultado la desmentiría y se informa del resultado aunque no guste. En el informe se habla de hipótesis «contrastada» o «no contrastada», nunca de hipótesis «demostrada» con una muestra pequeña.

ACTIVIDADES

4. Itxaso se plantea abrir una franja de recogida de kits para hogares de 20:30 a 21:30. Formula un objetivo general, tres objetivos específicos y una hipótesis contrastable para un estudio de dos semanas (del lunes 21/06 al viernes 02/07/2027) que le ayude a decidir.

5. Reescribe con población, variable y periodo, y di qué fallaba: a) «Confirmar que los hogares quieren reparto»; b) «Saber si el K05 sin gluten es bueno»; c) «Medir cuántos kits venderemos en primavera».

3.3 Exploratorio, descriptivo y causal

Explorar, describir y estimar un efecto son tres preguntas distintas y piden tres diseños distintos.

Tres diseños, tres conclusiones posibles. El **diseño** es la estrategia general del estudio: qué tipo de evidencia se va a reunir y qué permitirá concluir. No es la técnica (entrevista, encuesta, observación), aunque cada diseño se apoya en unas técnicas más que en otras.

El **exploratorio** es flexible y usa pocos casos para descubrir motivos y generar hipótesis; el **descriptivo** mide con un instrumento estructurado, de una vez (**transversal**) o a las mismas unidades en varios momentos (**longitudinal**, como un panel); el **causal** cambia una variable y la compara con un **grupo de control** (mejor, al azar) en el mismo periodo.

Un mismo estudio puede encadenar diseños: en el K03, primero el exploratorio (seis entrevistas, OE1 y OE4) y después el descriptivo (encuesta, OE2 y OE3).

Tabla 1 · Qué permite concluir cada diseño

Diseño	Pregunta típica	Permite concluir	No permite concluir
Exploratorio	¿Por qué no piden las oficinas?	Motivos y barreras posibles	Cuántas oficinas los tienen
Descriptivo	¿Qué proporción está interesada?	Proporciones con su margen	Que el interés cause compras
Causal	¿Aumenta las ventas el anuncio?	Efecto atribuible, si hay control	Efectos fuera de las condiciones probadas

El «antes y después» de Nora. Las ventas de kits de la semana del anuncio (lunes 01/02 a sábado 06/02) fueron mayores que las de la semana anterior. Eso es un **dato**; «el anuncio causó el aumento» es una **interpretación** que el dato no sostiene, porque faltan dos cosas: un **contrafactual** (qué habría pasado sin anuncio) y el control de otras causas.

Otras causas que cambiaron a la vez: calendario (primeros de mes, tras la cuesta de enero), **hechos puntuales** (el pedido grande de una oficina), **clima** (una semana de frío), **otras acciones** a la vez (promoción, precio, la web caída la semana anterior), **competencia** (el supermercado de enfrente) y **regresión a la media** (tras una semana anormalmente baja, la siguiente tiende a subir sin hacer nada).

La experimentación. Es la técnica del diseño causal: se manipula una variable y se compara con un **grupo de control**. Tres formas habituales: repetir el anuncio solo en zonas de reparto sorteadas y compararlas con el resto en las mismas semanas; una **prueba A/B**, con dos versiones (anuncio, precio, web) asignadas al azar; y el **test de mercado**, que lanza el producto en una zona y un periodo limitados (por ejemplo, el K03 solo en el polígono durante cuatro semanas) y mide ventas reales; sin zona de comparación, mide aceptación, no efecto. Un código de descuento exclusivo cuenta quién llegó por el anuncio, no si habría comprado igual.

ACTIVIDADES

6. Elige el diseño adecuado para cada encargo y justifícalo: a) Olvido quiere saber por qué algunas familias compran el K02 una vez y no repiten; b) Nora quiere estimar qué proporción de hogares de la zona de reparto compraría el K04 Cena lista una vez por semana; c) Unai quiere saber si el anuncio de febrero aumentó las ventas de kits.
7. Escribe cuatro explicaciones distintas del anuncio para el aumento de ventas de la semana del 01/02 y, para cada una, el dato interno que consultarías para comprobarla.

3.4 Método y fuentes

Primero se usa lo que ya existe y es pertinente; solo después se paga por datos nuevos.

Las fases y sus dependencias. Problema → diseño → fuentes → recogida → análisis → informe: el orden es fijo, pero se vuelve atrás si una fuente secundaria ya responde un objetivo (se rehace el diseño), si el pretest falla o no hay acceso (instrumento o muestra) o si en el análisis falta una variable.

Fuentes: dos criterios a la vez. Una fuente es **primaria** o **secundaria** (recogida para este estudio o para otro fin) e **interna** o **externa**: son dos clasificaciones distintas. Los pedidos de empresa son secundarios e internos; la encuesta a oficinas será primaria y externa.

Cómo se eligen. Para cada objetivo: ¿hay una fuente secundaria **pertinente** (misma población, variable y periodo), **fiable**, **representativa** y **actualizada**? ¿Cuánto cuesta y tarda? Una fuente barata que habla de otra población no ahorra nada: lleva a una conclusión equivocada. Y una primaria cara no se justifica si una secundaria pertinente ya responde: primero se agota lo que existe, después se paga por lo que falta.

Tabla 2 · Cuatro fuentes para el estudio del K03

Fuente	Tipo	Qué aporta	Límite y coste
A. Pedidos de empresa de La Despensa (enero de 2027)	Secundaria interna	Qué piden las oficinas que ya compran	Solo clientes; gratis
B. Listado de empresas del polígono de la asociación de empresarios	Secundaria externa	Marco del polígono y, por el censo de la cámara, del barrio	Fecha y altas sin comprobar; gratis
C. Seis entrevistas a responsables de oficinas	Primaria externa	Barreras, decisores, vocabulario	Pocos casos; 54,00 €
D. Encuesta a oficinas del barrio y del polígono	Primaria externa	Proporción interesada y condiciones	Necesita marco y pretest; 201,00 €

Orden: A y B (gratis, preparan el resto), C (redactar el cuestionario) y D. Ninguna secundaria responde OE2: ese hueco justifica pagar datos nuevos. Un informe gratuito de una plataforma nacional de kits a domicilio se descarta para OE2: habla de hogares de todo el país, no de oficinas del barrio.

Árbol del estudio: pregunta → objetivo → diseño → fuente → instrumento. La **técnica** es el modo de obtener los datos (entrevista, encuesta, observación); el **instrumento**, el documento que se usa (guion, cuestionario). Cada rama del árbol tiene que llegar hasta el instrumento: si no, es un objetivo que nadie medirá.

Tabla 3 · Árbol del estudio del K03

Objetivo	Diseño	Fuente	Instrumento
OE1 Barreras y motivos	Exploratorio	C. Entrevistas (6)	Guion de entrevista
OE2 Proporción interesada	Descriptivo	D. Encuesta	Cuestionario v0.1 → v0.2
OE3 Condiciones de compra	Descriptivo	D. Encuesta y A. Pedidos	Cuestionario y hoja de pedidos
OE4 Quién decide y paga	Exploratorio	C. Entrevistas y D. Encuesta	Guion y pregunta de clasificación

ACTIVIDADES

8. Ordena las fases del estudio del K03 (problema, diseño, fuentes, recogida, análisis e informe) y señala dos momentos en los que habría que volver a una fase anterior: qué lo provocaría y qué se cambiaría.

9. Para la decisión de la franja de recogida de 20:30 a 21:30 (actividad 4) dispones de cuatro fuentes: a) las hojas de recogida de la tienda de enero y febrero de 2027; b) la libreta de peticiones del mostrador; c) un estudio de horarios de compra de una asociación de consumidores de otra comunidad, de 2023; d) una encuesta breve a los titulares de la tarjeta. Clasifica cada una, valora su fiabilidad, representatividad y coste, y decide en qué orden las usarías antes de producir datos nuevos.

3.5 Plan, presupuesto y calendario

El plan convierte la pregunta en un trabajo con fechas, personas, dinero y riesgos que alguien puede aprobar.

El encargo de investigación. Antes del plan se fija por escrito el **encargo** (*briefing* del estudio): uso de la información, quién la recibe, qué necesita saber, fecha de decisión, presupuesto y límites. Es la referencia para saber después si el estudio ha cumplido y para decir qué queda fuera: aquí, las ventas futuras, el efecto del anuncio y las oficinas de fuera del barrio y del polígono.

Encargo del K03: **uso**, decidir si se ofrece el K03 a oficinas y en qué condiciones; **receptora**, Nora Castro, que decide; **necesidad**, interés declarado, condiciones de compra y quién decide en cada oficina; **fecha de decisión**, lunes 15/03/2027; **presupuesto y plazo**, 300,00 € y cuatro semanas (15/02 a 12/03).

Qué contiene el plan. Decisión y pregunta; objetivos e hipótesis; diseño; fuentes; técnicas e instrumentos; **muestra prevista** (población, marco, cuántas respuestas y cómo se eligen); **análisis** (qué tabla o cálculo responde a cada objetivo); presupuesto; calendario con responsables; y **riesgos** con su respuesta. Si falta una pieza, el plan no se puede aprobar ni, después, comprobar.

Calendario del estudio del K03

(días laborables, de lunes a viernes)

Semana	Tarea	Responsable	Entregable
1 · 15-19/02	Fuentes A y B; seis entrevistas (entre ellas, Mencía, de Albada)	Equipo de análisis	Notas codificadas
2 · 22-26/02	Cuestionario v0.1, pretest con cinco oficinas y v0.2	Equipo; Adrián prueba las rutas	Cuestionario v0.2
3 · 01-05/03	Recogida: oficinas sorteadas, enlace y llamadas	Unai (enlace); Adrián (registro de contactos)	Base de respuestas
4 · 08-12/03	Depuración, tablas, informe	Equipo de análisis	Informe EK-2027/IN
Lunes 15/03	Presentación y decisión	Nora	Decisión firmada



Presupuesto. Recoge todos los gastos reales del estudio, de la preparación al análisis. El tiempo del equipo de análisis y de la plantilla no se presupuesta en este caso (forma parte de su trabajo); la hoja de cálculo y el formulario en línea tampoco tienen coste añadido.

Presupuesto del estudio del K03 (precios sin IVA; costes de La Despensa)

Partida	Importe (€)
Seis entrevistas: una comida de muestra por persona (coste 9,00 €)	54,00
Material impreso: tarjetas con enlace y carteles	36,00
Pretest: cinco comidas de muestra (coste 9,00 €)	45,00
Agradecimiento: 60 degustaciones (coste 2,00 €)	120,00
Imprevistos (15 % del total)	45,00
Total	**300,00**

Directas: 255,00 €; imprevistos: 0,15 × 300,00 = 45,00 €. Las 60 degustaciones fijan la **muestra prevista**: hasta 60 respuestas de oficinas sorteadas del marco; solo así el margen sería de ±12,7 puntos al 95 % (p = q = 0,5; unidad 8).

Riesgos y respuestas. Un plan serio dice qué puede fallar antes de empezar:

- **Acceso**: quien contesta no siempre decide. Se pregunta por quien gestiona las compras y se anota su cargo.
- **No respuesta**: contestan sobre todo las oficinas que ya conocen La Despensa. Se registran contactos y respuestas, y se comparan clientes y no clientes.
- **Plazo**: si el pretest falla, la recogida se retrasa. Se reserva el viernes 26/02 para corregir.
- **Interpretación**: interés declarado no es compra. Si sale bien, se recomienda un piloto con pedidos reales, que mide lo que la encuesta no puede medir.

Ética: investigar no es vender. La **investigación comercial** es la obtención y el análisis sistemáticos de información para que una organización decida mejor; su finalidad es conocer, no convencer. Por eso un estudio no se usa para vender: si la «encuesta» acaba ofreciendo el K03 o pidiendo un pedido, ya es una acción comercial y hay que decirlo (ocultar el propósito comercial de una práctica es desleal: Ley 3/1991, de Competencia Desleal, art. 7.1).

El Código ICC/ESOMAR. La profesión se autorregula con el **Código Internacional ICC/ESOMAR** (edición de 2025; no es una ley, sino un código de conducta de la Cámara de Comercio Internacional y de ESOMAR). Sus cinco principios, resumidos: actuar de forma legal, honesta, transparente y veraz; tratar a las personas con cuidado y respeto, sin perjudicarlas; explicar con claridad cómo se recogen y usan sus datos, y protegerlos; no hacer nada que dañe la confianza del público en la investigación; y responder de todo el estudio, con cualquier técnica o tecnología. En el K03: no mezclar la consulta con una oferta. Consulta la versión vigente.

LA NORMA DICE

Reglamento (UE) 2016/679 (RGPD), de 27 de abril de 2016, art. 5.1: los datos personales serán «recogidos con fines determinados, explícitos y legítimos» (b) y «adecuados, pertinentes y limitados a lo necesario en relación con los fines» (c). El plan informa (art. 13) de para qué se recogen los contactos y no pide más de los necesarios. Consulta la versión vigente.

ACTIVIDADES

10. Pasa el encargo del K03 a una ficha y añade los límites del estudio y el entregable final (qué recibe Nora, cómo y cuándo).
11. Prepara el plan del estudio de la franja de 20:30 a 21:30 (actividad 4) con fases, instrumento, muestra prevista, análisis, recursos y riesgos: **120,00 €**, del 21/06 al 02/07/2027; Nora decide el lunes 05/07.
12. Unai propone 80 degustaciones. Con lo demás igual e imprevistos del 15 % del total, ¿cabe en 300,00 €? ¿Cuántas caben como máximo?

CASO RESUELTO · PASO A PASO

El plan del estudio del K03 para oficinas

Caso ficticio. Lunes 8 de febrero de 2027. Nora quiere ofrecer el K03 (20,00 € sin IVA) a las oficinas del barrio y del polígono. Tu equipo convierte su encargo en un plan aprobable.

Paso 1. Decisión, datos y alcance. La decisión es ofrecer o no el K03 a oficinas, y en qué condiciones, el lunes 15/03. Datos: pedidos de empresa de enero (solo clientes), listado del polígono y censo de la cámara (barrio). Restricciones: 300,00 € y cuatro semanas. Alcance: oficinas del barrio y del polígono, febrero y marzo de 2027; **no** mide ventas futuras ni el efecto del anuncio.

Paso 2. Procedimiento. La pregunta «¿demostrar que funcionará?» se reformula: «¿Qué proporción de oficinas declara que encargaría el K03 al menos una vez al mes a 20,00 € sin IVA, y en qué condiciones?».

Elemento	Decisión del plan
Objetivo general	Estimar interés declarado y condiciones de compra
Hipótesis	Al menos una de cada cuatro oficinas declara interés mensual
Diseño	Exploratorio (6 entrevistas) y descriptivo (encuesta)
Fuentes	Pedidos y listado (secundarias); entrevistas y encuesta (primarias)
Muestra prevista	Hasta 60 respuestas de oficinas del marco completo
Análisis	Proporción con su base; tablas por tipo de oficina
Presupuesto	54,00 + 36,00 + 45,00 + 120,00 + 45,00 = 300,00 €
Calendario	15/02 a 12/03; decisión el lunes 15/03

Unidades y condiciones: porcentaje **de oficinas que responden**, con su base (n); margen de error solo si se eligen al azar del marco.

Paso 3. Comprobación e interpretación. El presupuesto suma 300,00 € (imprevistos, 15 % del total); las semanas son de lunes a viernes; cada objetivo llega a su instrumento. El plan **permite concluir** qué proporción de las oficinas consultadas declara interés y en qué condiciones lo encargarían. **No permite concluir** cuántos kits se venderán, ni que el anuncio de febrero causara nada, ni nada sobre oficinas fuera del listado.

RESULTADO

Objetivo: estimar el interés declarado y las condiciones de compra del K03 en oficinas. Diseño: entrevistas exploratorias y encuesta descriptiva, con 300,00 € en cuatro semanas y decisión el 15/03. El efecto del anuncio queda fuera: pediría un diseño con control.

Comprueba que lo has entendido. a) ¿Por qué la hipótesis dice «declara»? b) ¿Qué cambiaría en el plan si Nora quisiera saber el efecto del anuncio? c) ¿Por qué las entrevistas van antes de la encuesta?

PRÁCTICA

El estudio de horarios de revisión de Ruedalia

Caso ficticio. Txema, dueño de Ruedalia, quiere ofrecer a los trabajadores del polígono **revisiones de bicicleta con cita** en franjas nuevas (por ejemplo, antes de las 9:00 o a la salida del turno). Ousmane cree que «a todos les viene bien a primera hora». Txema tiene **200,00 €** (entrevistas 60,00 €, impresos 40,00 €, incentivos 70,00 €, imprevistos 30,00 €) y **tres semanas**, del lunes 19/04 al viernes 07/05/2027.

Encargo. En parejas, aplica el procedimiento del caso resuelto: 1) separa el síntoma, la decisión y el problema de información, y reformula la idea de Ousmane como pregunta investigable; 2) escribe el objetivo general, dos específicos y una hipótesis contrastable; 3) elige diseño, fuentes (al menos una interna y una externa) y una muestra accesible; 4) reparte las tareas en las tres semanas con responsable; 5) comprueba que el presupuesto suma 200,00 € y di qué no podrá concluir el estudio.

Plantilla IC-P03 · Plan del estudio de horarios de revisión de Ruedalia

Campo	Tu respuesta
Datos y fuentes con unidad y periodo	
Procedimiento o cálculo	
Resultado y evidencia	
Decisión y limitaciones	

Lista de cotejo

☐ La pregunta admite una respuesta negativa.
☐ Población, variable y periodo en cada objetivo.
☐ Hipótesis con el resultado que la desmentiría.
☐ Diseño justificado (ninguna causa sin control).
☐ Fuentes secundarias antes que primarias.
☐ Calendario de lunes a viernes, dentro de las tres semanas.
☐ Presupuesto de 200,00 € comprobado.
☐ Responsable de cada tarea.

LECTURA Y DEBATE

El anuncio de febrero



Relato ficticio. El lunes 8 de febrero, Nora llegó a la trastienda con el móvil en la mano. «Mira las ventas de kits de la semana pasada: han subido. El anuncio de Unai ha funcionado. Esta semana lo repetimos y que Itxaso prepare más kits.» Unai sonrió: era su primer anuncio pagado.

Itxaso abrió la hoja de pedidos y separó las líneas una a una. «Una parte de la subida es un solo pedido: la jornada de formación de una oficina del barrio, que lo encargó por correo antes de que saliera el anuncio. Y el miércoles y el jueves hizo un frío de los que no se sale de casa: los repartos de tarde fueron el doble.» Adrián, que había anotado los repartos, lo confirmó con su libreta.

«Entonces, ¿el anuncio no sirvió?», preguntó Nora. «No lo sabemos», dijo Itxaso. «Puede que sí y puede que no. Con una semana antes y una después no se puede separar.» Unai propuso entonces una prueba: repetir el anuncio solo para la mitad de las calles de la zona de reparto, elegidas por sorteo, y comparar las dos mitades durante tres semanas. Nora aceptó, pero con una condición: no preparar más kits hasta ver la comparación.

Preguntas. 1) ¿Qué afirmación del relato necesitaba una comprobación antes de convertirse en decisión? 2) ¿Qué información pidió Itxaso primero y por qué cambió la respuesta? 3) ¿Qué actuación provisional se acordó para aprender sin ocultar la incertidumbre? ¿Propondrías otra?

APRENDE CON LA IA

Unai pregunta a un asistente de IA cómo demostrar a Nora que su anuncio funcionó. Esta es la respuesta:

RESPUESTA DE LA IA

«Es sencillo: si las ventas de la semana del anuncio fueron mayores que las de la semana anterior, eso prueba que el anuncio causó la subida. Una subida de ventas posterior a la campaña prueba causalidad. Para que el informe sea más convincente, calcula el porcentaje de aumento y preséntalo como el efecto del anuncio.»

Tu tarea. Localiza el error de razonamiento, contrástalo con lo que has aprendido en el apartado 3.3 (contrafactual, otras causas, grupo de control) y escribe una respuesta corregida para Unai que diga qué se puede afirmar con esos datos, qué no y qué diseño haría falta. Conserva la prueba que justifica tu corrección (por ejemplo, los hechos de esa semana que también pudieron influir).

Ojo con la IA. Un asistente puede tomar «después de» por «a causa de»: pídele qué otras explicaciones caben. **Sin asistente:** trabaja con este recuadro.

ACTIVIDADES FINALES

Comprueba

13. Resume con tus palabras la diferencia entre un diseño descriptivo y uno causal, y aplícala al caso resuelto: ¿por qué el estudio del K03 no es causal y no necesita serlo?
14. Explica qué tiene que tener una hipótesis para ser contrastable y comprueba si lo cumple la hipótesis del caso resuelto.
15. Di qué fuentes del caso son primarias o secundarias e internas o externas, y por qué las secundarias se usan primero.
16. Explica qué es un test de mercado y en qué se diferencia de la encuesta del K03: qué mide cada uno y qué decisión apoya.

Aplica

17. Repite la práctica de Ruedalia cambiando una condición: Txema solo tiene **dos semanas** (del lunes 19/04 al viernes 30/04/2027) y el mismo presupuesto. Señala qué partes del plan cambian (calendario, muestra, técnicas) y cuáles se mantienen (pregunta, objetivos, hipótesis).
18. Nora te pide añadir al estudio del K03 la pregunta «¿cuánto subieron las ventas por el anuncio?». Explica por qué no cabe en este plan y propón, en cinco líneas, un diseño aparte para responderla.
19. Si las comidas de muestra del K03 costaran 10,00 € en lugar de 9,00 €, con las mismas partidas y los imprevistos al 15 % del total, ¿cuánto costaría el estudio? ¿Qué partida recortarías para seguir en 300,00 € y qué riesgo añade el recorte?

Investiga

20. Localiza una fuente primaria o documentación técnica útil para el plan, el presupuesto o el calendario de un estudio (por ejemplo, la guía o el código de buenas prácticas de una asociación profesional de investigación de mercados, o una plantilla de plan publicada por un organismo público). Registra la fecha de consulta, el ámbito, la afirmación que te sirve para tu plan y una limitación.

TU INVESTIGACIÓN AVANZA · PASO 3 · I03 · PLAN DE INVESTIGACIÓN

Incorpora la entrega **I03 Plan de investigación** a tu dosier: para el negocio que elegiste en el paso 1, escribe la decisión, la pregunta investigable, el objetivo general con tres específicos, una hipótesis contrastable, el diseño, las fuentes por orden de uso, la muestra prevista, el análisis, el calendario con responsables, el presupuesto y los riesgos. Usa el caso resuelto como modelo.

Comprueba tu dosier: cada objetivo llega hasta su instrumento; el presupuesto suma; incluyes la evidencia (hoja de cálculo del presupuesto y calendario), la decisión que apoya el plan, el responsable y la fecha de revisión. Si corriges una versión, conserva el motivo del cambio.

☐ Hecho ☐ Revisado ☐ Guardado

OJO PROFESIONAL

Documento: borrador del plan del estudio del K03 (EK-2027/PL-v0.1)

Documento ficticio de La Despensa de Nora, S.L., redactado el martes 09/02/2027 para que Nora lo apruebe. Contiene **cinco errores**.

N.º	Apartado	Contenido del borrador
1	Objetivo general	Demostrar que el K03 gustará a las oficinas del barrio y del polígono
2	Población	Oficinas del barrio y del polígono, febrero y marzo de 2027
3	Antecedente	La subida de ventas tras el anuncio de febrero fue un experimento controlado: el anuncio funciona
4	Diseño	Seis entrevistas exploratorias y una encuesta descriptiva
5	Muestra	Las oficinas que ya compran kits, porque responden más
6	Presupuesto	Recogida: impresos 36,00 € y degustaciones 120,00 € = 156,00 €; la preparación y el análisis «no cuentan»
7	Calendario	Del lunes 15/02 al viernes 12/03/2027
8	Decisión	Viernes 05/03/2027, con los primeros resultados

Encargo. Señala los cinco errores, explica qué falla en cada uno con lo que has aprendido en la unidad y reescribe las líneas afectadas para que Nora pueda aprobar el plan. **Para pensar:** ¿qué apartados faltan por completo en el borrador?

N.º	Línea	Por qué es un error	Corrección
1			
2			
3			
4			
5			

AUTOEVALUACIÓN

1. Las ventas internas de kits permiten saber...

a) qué compran los clientes actuales, no qué piensan las oficinas que no compran.
b) qué proporción de oficinas está interesada en el K03.
c) si el anuncio causó el aumento.

2. Un objetivo específico bien formulado es...

a) demostrar que el K03 gustará.
b) estimar la proporción de oficinas del barrio y del polígono que declara que encargaría el K03 al mes (febrero-marzo de 2027).
c) saber qué opinan los clientes.

3. Para evaluar el efecto de una campaña sobre las ventas hace falta...

a) comparar la semana anterior con la posterior.
b) un diseño con grupo de control en el mismo periodo, asignado al azar.
c) seis entrevistas en profundidad.

4. En el estudio del K03, la encuesta va después de las entrevistas porque...

a) las entrevistas aportan barreras y vocabulario para redactar el cuestionario.
b) la encuesta es más barata.
c) las entrevistas estiman la proporción de interesados.

5. Un encargo de investigación completo incluye...

a) solo el presupuesto y el calendario.
b) el uso, la receptora, la necesidad de información y la fecha de decisión.
c) el resultado esperado del estudio.

6. El presupuesto del caso suma 255,00 € de partidas directas. Si los imprevistos son el 15 % del total, el total es...

a) 293,25 €.
b) 300,00 €.
c) 270,00 €.

7. Ousmane cree que «a todos les viene bien a primera hora». La pregunta investigable es...

a) ¿cómo demostramos que la primera hora es la mejor?
b) ¿qué franjas prefieren los trabajadores del polígono y con qué frecuencia las usarían (abril-mayo de 2027)?
c) ¿cuántas revisiones haremos este año?

8. «Una subida de ventas posterior a la campaña prueba causalidad.» Es falso porque...

a) las subidas nunca tienen causa.
b) falta el contrafactual: otras causas (calendario, clima, pedidos puntuales) pudieron cambiar a la vez.
c) solo un porcentaje de aumento prueba causalidad.

9. En el Ojo profesional, la línea 6 falla porque...

a) deja fuera la preparación (entrevistas y pretest), los imprevistos y la fase de análisis.
b) los impresos cuestan más.
c) debería incluir el IVA.

10. En tu entrega I03, ¿qué dato te obligaría a revisar el plan antes de recoger datos?

a) Que una fuente secundaria ya responda un objetivo.
b) Que el cuestionario sea largo de leer.
c) Que la hipótesis te parezca probable.

Ya sé...

- ☐ explicar las distinciones de los cinco apartados.
- ☐ convertir un encargo en un plan que se puede aprobar.
- ☐ transferir el procedimiento a otro negocio (Ruedalia).
- ☐ detectar una conclusión no sustentada («después de, luego por»).
- ☐ entregar el plan I03 como evidencia revisable.



UNIDAD 4

Fuentes secundarias y suficiencia

Bloque 2 · Fuentes e instrumentos

EN ESTA UNIDAD APRENDERÁS A…

- Planificar una búsqueda online y offline con buscadores, operadores avanzados y criterios de selección, y registrarla.
- Valorar por separado la fiabilidad, la cobertura, la actualidad y las condiciones de uso de una fuente.
- Normalizar datos con sus metadatos, leer una serie temporal y deflactar importes con un índice de precios.
- Decidir, objetivo por objetivo, si la información disponible basta y qué falta por preguntar.
- Archivar las fuentes con versión, responsable y fecha para que otro equipo pueda recuperarlas.

SITUACIÓN DE PARTIDA

Lunes 22 de febrero de 2027. Unai llega con un informe gratuito que ha encontrado en internet: «El 38 % de los hogares compraría kits de comida preparada». Nora hace la cuenta en voz alta: «Si en el barrio hay unos 6.000 hogares, son 2.280 hogares. No hace falta encuesta».

- Adrián lee la letra pequeña: la encuesta es de otra ciudad, de 2022, la contestó quien quiso y cuenta como «hogar» a cada persona que responde.
- Sobre la mesa hay tres fichas: **A**, el padrón; **B**, ese informe; **C**, las ventas de la prueba de kits.

Nora te encarga el **inventario de fuentes y lagunas** del estudio de kits para hogares, para el viernes 26/02, antes de pagar ningún dato nuevo.

PARA EMPEZAR

1. ¿Qué quiere decidir Nora con el 38 % y qué tendría que comprobar antes?
2. ¿Qué sabemos de verdad sobre los hogares del barrio y qué estamos suponiendo?
3. ¿Qué dato del informe permitiría revisar la cuenta de Nora?
4. ¿Por qué no se pueden multiplicar personas por un porcentaje de «hogares»?



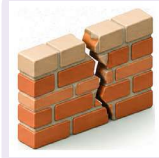
Búsqueda 4.1



Fiabilidad 4.2



Metadatos 4.3



Lagunas 4.4



Archivo 4.5

4.1 Búsqueda online y offline

Se busca con una pregunta delante: primero se decide qué dato hace falta y después dónde puede estar.

Lo que ya sabes. Distingues fuentes primarias y secundarias, internas y externas, y sabes citar una fuente con su fecha de consulta. Ahora buscarás con método, para un objetivo concreto del plan.

Plan de búsqueda. Antes de abrir un buscador se escribe, para cada objetivo, qué variable, de qué población y de qué periodo se necesita. Del objetivo salen los **términos** (y sus sinónimos: «comida preparada», «platos cocinados», «comida para llevar») y los **lugares** donde buscar.

Fuentes online. Organismos estadísticos (el INE, autoridad estadística nacional, publica resultados agregados de sus operaciones), portales de datos abiertos de las administraciones, ayuntamientos, cámaras de comercio, asociaciones sectoriales y de consumidores, y publicaciones especializadas.

Fuentes offline. Archivo interno (tiques, hojas de pedidos, libreta de peticiones), publicaciones impresas, bibliotecas y documentación que se pide a un organismo. Lo offline también hay que fecharlo y registrarlo, porque nadie lo encontrará en un buscador.



Motores de búsqueda. Un **buscador generalista** rastrea toda la web y ordena los resultados por relevancia, popularidad y anuncios: el primer resultado no es el mejor, sino el mejor posicionado. Los **buscadores internos** de un portal oficial o de una base de datos buscan solo dentro de ella y permiten filtrar por tema, territorio y periodo. Un **catálogo de biblioteca** localiza documentos impresos y digitales.

Operadores de búsqueda avanzada. Acotan la consulta. Funcionan en los buscadores generalistas más usados, pero cada uno tiene los suyos: compruébalos en su ayuda.

Operador	Qué hace	Ejemplo
"..."	Busca la frase exacta	"comida preparada" hogares
-	Excluye un término	kits comida -receta

Operador	Qué hace	Ejemplo
OR	Acepta cualquiera de los dos	padrón OR población
site:	Limita a un dominio	padrón site:ine.es
filetype:	Limita al tipo de archivo	encuesta hogares filetype:pdf

Criterios de selección. De cada resultado se mira: quién lo publica y con qué propósito (informar, vender, opinar); si explica el método (población, muestra, fechas, definiciones); el ámbito y el periodo; la fecha de actualización; las condiciones de uso; y el coste. Un resultado sin autor ni método no se descarta para siempre, pero no sirve como dato. Entre dos fuentes parecidas se prefiere la más cercana al origen y con el método publicado, aunque sea menos vistosa.

Registro de búsqueda.

Cada búsqueda deja una línea: fecha, buscador o lugar, consulta exacta, resultados revisados, fuente elegida y motivo. Así otro equipo puede repetirla. Ejemplo: «22/02/2027 · buscador general · "comida preparada" hogares filetype:pdf · 30 resultados revisados · elegido: informe de consumidores (ficha B) · límite: otra ciudad, 2022, «hogar» = persona».

Ir al origen. Muchas páginas repiten una cifra que sacaron de otra. Antes de usarla se sigue la cadena hasta el **origen primario** (quien recogió el dato) y se anotan su definición y su método, si están publicados. Si la cadena se corta, la cifra se registra como «origen no localizado» y solo vale como pista. Una cifra redonda que aparece en muchas webs sin fecha ni autor suele ser la misma cifra copiada, no varias confirmaciones.

Errores habituales. Quedarse con la primera página de resultados; confundir la fecha de publicación con la fecha de los datos; tomar un resumen de prensa por el informe; y buscar solo lo que confirma la idea de partida: el sesgo de confirmación de la unidad anterior, ahora en la búsqueda.



CLAVE · BUSCAR NO ES ACUMULAR

Una búsqueda empieza en el objetivo y termina en un registro. Se guarda lo que responde a una variable, una población y un periodo del plan, con su origen, su definición y su fecha; lo demás, aunque sea interesante, se queda fuera.

ACTIVIDADES

1. Para el objetivo «cuántos hogares hay en el barrio», recupera una fuente online oficial y un documento offline (autorizado o preparado por tu profesorado). Registra de cada uno la ubicación, la fecha de consulta, el alcance y el dato extraído.
2. Escribe tres consultas con operadores para buscar estudios sobre compra de comida preparada por hogares: una con frase exacta y exclusión, otra limitada a un dominio oficial y otra a documentos PDF. Explica qué resultados quita cada operador.

4.2 Fiabilidad y representatividad

Una fuente prestigiosa de otra población no se vuelve representativa del barrio por ser prestigiosa.

Cuatro preguntas por separado. Una fuente puede aprobar una y suspender otra (el padrón es muy fiable y, aun así, no habla de compras), así que no se le pone una nota única:

- **Fiabilidad**: ¿quién la elabora, con qué método documentado y con qué control de errores? ¿Se puede comprobar?
- **Cobertura** (o representatividad): ¿la población, el territorio y la definición coinciden con los del objetivo?
- **Actualidad**: ¿de cuándo son los datos (no la publicación)? ¿Ha cambiado algo desde entonces?
- **Condiciones de uso y seguridad**: ¿se puede reutilizar y citar? ¿Contiene datos personales? ¿El archivo es íntegro?

Representatividad. Una muestra representa a una población si el marco la cubre, se eligió con un diseño apropiado (mejor, al azar) y se trata la no respuesta: poder participar no basta. En una encuesta **voluntaria** en línea responde quien quiere, sobre todo los más interesados, y ni 400 respuestas ni 4.000 lo corrigen. El tamaño reduce el azar, no el sesgo. Por eso la ficha B, con 412 respuestas, es menos representativa que una con 100 elegidas al azar del barrio.

Dato, definición y fecha. Dos fuentes que usan la misma palabra pueden medir cosas distintas. En el uso habitual, el **hogar** es la persona o grupo que vive en la misma vivienda; en la ficha B, «hogar» es cada persona que responde. El 38 % de la ficha B no es una proporción de hogares en ese sentido: si en una casa contestaron dos personas, esa casa cuenta dos veces.

El padrón. La **ficha A** procede del padrón municipal, el registro de vecinos que gestiona cada ayuntamiento. Es un registro administrativo, no una encuesta: cuenta **personas** empadronadas, no hogares, y no dice nada de lo que compran. Sirve para el contexto de población.

También las fuentes internas. La ficha C es fiable para lo que registra (kits cobrados en caja), pero su cobertura tiene un hueco: 213 de los 648 kits se vendieron sin tarjeta y no se sabe a quién.

LA NORMA DICE

Ley 7/1985, de 2 de abril, reguladora de las Bases del Régimen Local, art. 16.1: «El Padrón municipal es el registro administrativo donde constan los vecinos de un municipio». Según su art. 17.3, los ayuntamientos remiten sus datos al INE para coordinar los padrones de todos los municipios. Consulta la versión vigente.

Opiniones y reseñas. Las opiniones de una red social o las reseñas de un comercio son fuentes secundarias externas con dos problemas: quien escribe se elige solo y no siempre se sabe si compró. La Ley 3/1991, de Competencia Desleal, considera desleal presentar reseñas como de clientes reales sin comprobarlo y encargar reseñas falsas (art. 27.7 y 27.8): un motivo más para tratarlas como indicio, nunca como medida.

Aceptar, limitar o descartar. Tras las cuatro preguntas, cada fuente recibe una decisión por objetivo: **aceptar** (responde con su límite escrito), **limitar** (solo como contexto o hipótesis) o **descartar** (llevaría a una conclusión equivocada). La decisión se anota junto a la fuente, con el criterio que la motivó, para que quien lea el inventario no tenga que repetir la auditoría.

ACTIVIDADES

3. Audita las fichas A, B y C de la situación de partida con las cuatro preguntas y decide, para el objetivo «estimar qué proporción de hogares del barrio compraría kits», si aceptas, limitas o descartas cada una.

4. Unai propone usar las 25 reseñas del perfil del supermercado de enfrente para saber qué opinan los vecinos de su comida preparada. Valora la fuente con las cuatro preguntas y di qué uso le darías.

4.3 Organización y metadatos

Los datos se normalizan sin borrar el original: unidad, periodo y definición viajan con cada cifra.

Tabla normalizada.

Antes de comparar, cada dato pasa a una tabla con las mismas columnas: variable, valor, unidad, periodo, población, fuente y limitación. Lo que no cabe en una columna se escribe en la limitación.

Tabla 1 · Fichas del inventario, normalizadas

Ficha	Variable y valor	Unidad	Periodo	Población	Limitación
A	Empadronados: 14.520	Personas	01/01/2026	Secciones del barrio	No da hogares ni compra
B	Compraría kits: 38 %	% de quienes responden (412)	2022	Otra ciudad, voluntarios	«Hogar» = persona que responde
C	Kits vendidos: 648	Kits	11/01-20/02/2027	Compradores de K01 y K02	213 kits sin cliente identificado

Metadatos. Son los datos que describen el dato: definición de cada variable, unidad, fuente con su enlace o ubicación, fecha de los datos y de la consulta, versión del archivo, transformaciones hechas y responsable. Sin ellos, dentro de seis meses nadie sabrá si «648» eran kits, tiques o euros, ni de qué semanas.

No mezclar unidades. Personas y hogares no se suman ni se dividen sin decirlo. Con 14.520 personas empadronadas y una estimación de unos 6.000 hogares, salen 14.520 ÷ 6.000 = **2,42 personas por hogar**: un dato derivado, que se anota con sus fuentes y fechas (padrón: 01/01/2026; hogares: sin fuente ni fecha).

Series temporales. Una **serie temporal** es una variable medida a intervalos regulares (por semanas, meses o años). Se descompone en cuatro componentes: **tendencia** (la dirección a largo plazo), **estacionalidad** (lo que se repite cada semana, mes o año, como las meriendas del sábado), **ciclo** (subidas y bajadas de varios años, ligadas a la economía) e **irregular** (lo imprevisto: un pedido grande, una semana de frío).

Tabla 2 · Kits vendidos por semana en la prueba (ficha C) y media móvil de tres semanas

Semana (lunes a sábado)	11-16/01	18-23/01	25-30/01	01-06/02	08-13/02	15-20/02
Kits	92	126	98	124	100	108
Media móvil (3)	—	105,33	116,00	107,33	110,67	—

La **media móvil** de tres semanas centrada promedia cada semana con la anterior y la siguiente: (92 + 126 + 98) ÷ 3 = 105,33. Suaviza lo irregular y deja ver la tendencia. Aquí la serie suavizada oscila alrededor de 108 kits (648 ÷ 6); las semanas de 126 y 124 tienen explicación propia (la prueba del K02, del 18 al 23/01, y un pedido de formación más el anuncio, sin separar, del 01 al 06/02). Seis semanas no bastan para ver una tendencia a largo plazo ni estacionalidad: haría falta un año de datos, mejor dos.

Deflactar con un índice de precios. Un importe en euros de distintos años mezcla dos cosas: lo que cambió la cantidad y lo que cambiaron los precios. **Deflactar** es pasar los importes a euros de un mismo año dividiendo por un índice de precios: valor real = valor nominal ÷ índice × 100. El índice habitual es el **IPC** (índice de precios de consumo) del INE (general, base 2025 = 100: 94,8 en 2023; consulta del 29/10/2026). Aquí, un índice de los kits ficticio.

Tabla 3 · Tique medio de la tienda (sin IVA) en euros corrientes y en euros de 2023

Año	2023	2024	2025	2026
Tique medio nominal (€)	12,80	13,30	13,70	14,00
Índice de kits (2023 = 100)	100,0	102,8	105,6	108,4
Tique medio real (€ de 2023)	12,80	12,94	12,97	12,92

De 2023 a 2026 el tique nominal sube un 9,4 %, pero en euros de 2023 solo un 0,9 %: casi toda la subida es de precios; no se mide la cantidad comprada. El cálculo de 2026: 14,00 ÷ 108,4 × 100 = 12,92 €; variación nominal (14,00 − 12,80) ÷ 12,80 = 9,4 %; variación real (12,92 − 12,80) ÷ 12,80 = 0,9 %.

Qué índice elegir. El IPC general sirve para deflactar un gasto de consumo variado, como el tique medio. Si se estudia un producto concreto, el INE publica también índices por grupos (alimentos, por ejemplo), que se ajustan mejor. Se anota en los metadatos cuál, con su base y fecha: otro índice cambia el resultado. Y nunca se deflacta una cantidad física (kits, tiques): solo importes en euros.

CLAVE · EL ORIGINAL NO SE TOCA

Se trabaja sobre una copia, y cada transformación (cambio de unidad, media móvil, deflactado) se anota en los metadatos con su fórmula. Así se puede volver al dato bruto, repetir el cálculo y comprobar que sale lo mismo.

ACTIVIDADES

5. Construye la tabla normalizada de tu inventario con las columnas variable, valor, unidad, periodo, población, fuente y limitación, e incluye las fichas A, B y C y el dato de personas por hogar con su origen.
6. En 2023, los vecinos de diario gastaban de media 8,40 € por visita; en 2026, 9,00 €. Con el índice de la Tabla 3, expresa el gasto de 2026 en euros de 2023. ¿Gastan más o menos en términos reales? ¿Qué porcentaje?
7. Las reservas de bizcocho para el sábado de seis semanas seguidas fueron 34, 40, 37, 45, 39 y 42. Calcula la media móvil de tres semanas centrada, describe la tendencia y di qué componente de la serie no puedes ver con estos datos.

4.4 Suficiencia y lagunas

La suficiencia se decide objetivo por objetivo, no por el número de páginas recopiladas.

Suficiencia cuantitativa y cualitativa. Una fuente es **suficiente** para un objetivo si da la variable, la población y el periodo que pide, con una precisión que permita decidir (análisis cuantitativo), y si es útil, fiable y representativa (análisis cualitativo). Diez informes sobre «hogares de España» no responden una pregunta sobre los hogares del barrio, por muchas páginas que sumen.

La matriz objetivo-evidencia-suficiencia. Es el recurso de esta unidad: una fila por objetivo, con la evidencia disponible, su límite y un veredicto: **cubierto**, **parcial** o **sin evidencia**. Lo que queda parcial o sin evidencia son las **lagunas**: las preguntas que justifican pagar información primaria. La matriz se rellena antes de diseñar el cuestionario y se revisa cada vez que entra una fuente nueva.

Cuándo basta. Un objetivo está cubierto si se cumplen cuatro condiciones: la fuente mide la variable del objetivo con su misma definición; habla de la misma población y de un periodo que sigue valiendo; su precisión permite elegir entre las opciones de la decisión; y el coste de equivocarse con ella es menor que el de pagar datos nuevos. Si falla una, el veredicto es parcial; si fallan las dos primeras, no hay evidencia; en otros casos, parcial si mide la variable y la población, y sin evidencia si no.

Tabla 4 · Matriz del estudio de kits para hogares (22/02/2027)

Objetivo	Evidencia disponible	Veredicto
O1. Hogares del barrio	A: 14.520 personas; estimación de unos 6.000 hogares sin fuente documentada	Parcial
O2. Proporción de hogares que compraría kits	B: otra ciudad, 2022, voluntaria, otra definición	Sin evidencia
O3. Perfil y frecuencia de los compradores actuales	C: 435 kits de 107 titulares (4,07 por titular); 213 sin identificar	Parcial
O4. Servicio de pedido preferido	Libreta de peticiones: peticiones sueltas de reparto por la tarde	Parcial (solo indicio)

Cómo se lee la matriz. O1 es parcial: el padrón da personas y los 6.000 hogares no tienen fuente. O2 no tiene evidencia: la ficha B ni mide hogares ni habla del barrio. O3 es parcial por los 213 kits sin identificar. O4 solo tiene peticiones espontáneas: un tema, no cuántos lo quieren.

De la laguna a la técnica primaria. Cada laguna se convierte en una pregunta pendiente con su técnica, su coste y su límite: O2 pide una encuesta a una muestra de hogares del barrio (unidades 6 a 8); O4, la misma encuesta o unas entrevistas exploratorias; O1, preguntar al ayuntamiento si publica hogares por sección antes de estimarlos. No se pregunta en la encuesta lo que ya está cubierto: cada pregunta de más alarga el cuestionario y baja la respuesta.

PARA SABER MÁS

El INE ofrece en su web un buscador de estadísticas por tema y un calendario de publicaciones. Conocer cuándo se actualiza una operación evita usar una cifra que está a punto de cambiar.

ACTIVIDADES

8. Nora estudia lanzar el **K04 Cena lista** (28,00 € sin IVA) para hogares con poco tiempo. Objetivos: a) cuántos hogares del barrio tienen a todas sus personas adultas trabajando fuera; b) qué proporción compraría el K04 una vez por semana; c) a qué precio; d) en qué franja de recogida. Dispones de la ficha A, la ficha C y la libreta de peticiones (cinco peticiones de «cena hecha» en enero). Marca cada objetivo como cubierto, parcial o sin evidencia y justifícalo.

9. Formula dos preguntas pendientes de la actividad 8 y elige para cada una la técnica primaria, su coste aproximado y su limitación.

4.5 Archivo y recuperación

Un dato que nadie puede volver a encontrar, o que nadie sabe de qué versión es, ya no es evidencia.

Estructura del archivo. Una carpeta por proyecto, con subcarpetas para **brutos** (las fuentes tal como llegaron, sin tocar), **trabajo** (copias normalizadas) y **resultados**. El nombre de cada archivo sigue el patrón `AAAA-MM-DD_tema_estado_vN`, por ejemplo `2027-02-22_padron_bruto_v1` y `2027-02-24_inventario_normalizado_v2`.

Ficha de fuente.

Cada fuente archivada lleva su ficha de metadatos: título, organismo, enlace o ubicación, fecha de los datos, fecha de consulta, condiciones de uso, versión, responsable y uso en el estudio.

Versiones y registro de cambios. Cuando una fuente se actualiza (el padrón de 2027 sustituye al de 2026), no se sobrescribe: se guarda la nueva versión con su fecha, se conserva la anterior y se anota en el **registro de cambios** qué cambió, quién, cuándo y por qué. Si un informe cita «14.520 personas», tiene que poder enlazar con la versión de la que salió esa cifra.

Acceso y seguridad. Los brutos se guardan en solo lectura; editan los responsables del estudio y Nora consulta.

Hay una copia de seguridad en otro soporte y se prueba la restauración. El bruto con datos personales (ficha C, titulares) se guarda cifrado o con acceso muy limitado; las copias de trabajo, seudonimizadas. Y no se guardan más tiempo del necesario: el RGPD pide que los datos que identifican a una persona se mantengan solo mientras lo exija el fin para el que se recogieron (art. 5.1 e).

Recuperación. La prueba de un buen archivo: otro equipo, sin preguntar, localiza el dato, lee su definición y reconstruye cada transformación.

LA NORMA DICE

Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público, art. 8: la reutilización puede estar sometida, entre otras, a estas condiciones: «a) Que el contenido de la información, incluyendo sus metadatos, no sea alterado», «c) Que se cite la fuente» y «d) Que se mencione la fecha de la última actualización». Consulta la versión vigente.

ACTIVIDADES

10. En una hoja de cálculo, guarda la ficha A con sus metadatos como `2027-02-22_padron_bruto_v1` en solo lectura; haz una copia de trabajo, cambia la unidad de una columna y anota el cambio en el registro. Después recupera la versión original y comprueba que no ha cambiado.

11. Diseña la ficha de fuente de la ficha B con todos los campos del apartado y escribe la línea del registro de cambios que anotarías si el informe publicara una edición de 2027.

CASO RESUELTO · PASO A PASO

El inventario de fuentes del estudio de kits

Caso ficticio. Lunes 22 de febrero de 2027. Nora quiere decidir si basta con la información que ya existe para dimensionar la línea de kits para hogares o si hay que pagar una encuesta. Tu equipo audita las tres fichas.

Paso 1. Decisión, datos y alcance. Decisión: encargar o no una encuesta a hogares antes de ampliar la producción de kits. Datos: A (padrón, 01/01/2026: 14.520 personas), B (informe de otra ciudad, 2022: 412 respuestas voluntarias, «el 38 % de los hogares compraría kits») y C (ventas de la prueba, 11/01-20/02/2027: 648 kits). Alcance: hogares del barrio en 2027.

Paso 2. Procedimiento. Cada ficha se valora por separado y se decide según su uso: objetivo principal (O2, proporción de hogares que compraría kits), contexto o clientela actual.

Ficha	Fiabilidad	Cobertura	Actualidad	Decisión
A · Padrón	Alta: registro oficial	Personas, no hogares; no mide compra	2026	Limitar a contexto
B · Informe	Baja: método sin control	Otra ciudad, voluntaria; «hogar» = quien responde	2022	Limitar a hipótesis
C · Ventas	Alta: registro propio	Solo compradores actuales	2027	Aceptar para clientela actual

Unidades: A en personas, B en % de quienes responden, C en kits. La cuenta de Nora (0,38 × 6.000 = 2.280 hogares) aplica un porcentaje de otra población y otra definición a una estimación sin fuente: no es un dato.

Paso 3. Comprobación e interpretación. C es coherente: 435 kits de 107 titulares (435 ÷ 107 = 4,07 kits por titular) más 213 sin identificar suman 648 (108 por semana). Ninguna ficha cubre O2: es la laguna que justifica la encuesta. En la matriz: O1 parcial (personas, no hogares), O2 sin evidencia, O3 parcial (213 kits sin identificar) y O4 parcial (solo peticiones sueltas de la libreta). **Permite concluir** cuántas personas están empadronadas en el barrio y cómo compran los clientes con tarjeta. **No permite concluir** cuántos hogares comprarían kits.

RESULTADO

A contextualiza la población, no la demanda; B limita la extrapolación (otra ciudad, otra fecha, otra definición, respuesta voluntaria); C describe la clientela actual. Antes de recoger datos se registran los huecos: O2 sin evidencia; O1, O3 y O4, parciales.

Comprueba que lo has entendido. a) ¿Por qué el padrón es fiable y aun así no responde O2? b) ¿Qué tendría que cambiar en la ficha B para poder aceptarla? c) ¿Por qué se registran los huecos antes de diseñar la encuesta?

PRÁCTICA

Las fuentes del servicio a empresas de Ruedalia

Caso ficticio. Txema quiere decidir si ofrece a las empresas del polígono **revisiones de bicicleta en sus instalaciones**. Ousmane ha reunido dos fuentes:

- **Informe nacional de 2026** de una asociación del sector de la bicicleta: «el 12 % de las personas ocupadas va al trabajo en bicicleta al menos una vez por semana». Encuesta en línea a 2.000 personas de todo el país, con cuotas por edad y sexo.
- **40 opiniones** publicadas en un grupo de una red social de vecinos del polígono (enero y febrero de 2027): 28 piden más aparcamientos para bicis y 9 mencionan averías.

Txema calcula: «El polígono tiene 1.400 trabajadores: el 12 % son 168 ciclistas. Y el 70 % de las opiniones (28 de 40) quiere el servicio».

Encargo. En parejas, aplica el procedimiento del caso resuelto: 1) escribe la decisión, los datos y el alcance; 2) valora cada fuente por fiabilidad, cobertura, actualidad y condiciones de uso, y decide aceptar, limitar o descartar; 3) normaliza las dos cifras de Txema (unidad, población, periodo) y di qué falla en cada cálculo; 4) completa la matriz objetivo-evidencia-suficiencia para dos objetivos (ciclistas del polígono y empresas interesadas); 5) formula la laguna principal y la técnica primaria que la cubriría.

Plantilla IC-P04 · Inventario de fuentes de Ruedalia

Campo	Tu respuesta
Datos y fuentes con unidad y periodo	
Procedimiento o cálculo	
Resultado y evidencia	
Decisión y limitaciones	

Lista de cotejo

☐ Cada fuente con unidad, población y periodo.
☐ Las cuatro preguntas por separado.
☐ Las opiniones voluntarias no se tratan como muestra.
☐ El ámbito nacional no se aplica sin más al polígono.
☐ Matriz con cubierto, parcial o sin evidencia.
☐ Laguna convertida en pregunta con técnica y coste.
☐ Fecha de consulta de cada fuente.

LECTURA Y DEBATE

Otra definición de hogar



Relato ficticio. El martes, Nora ya había pedido a Itxaso que calculara cuántos envases harían falta para «unos 2.280 hogares». Itxaso no se puso a calcular: «Antes de comprar envases, ¿de dónde sale esa cifra?».

Adrián había impreso el informe y marcado una línea de la última página con un círculo: «Hogar: persona que responde a la encuesta». Unai llamó, con el altavoz puesto, a la asociación de consumidores que lo había publicado. Le atendió Laia, la técnica que lo coordinó. «Lo hicimos con un formulario abierto en nuestra web durante un mes. Si en una casa contestaban dos personas, contaban como dos hogares. Sirve para saber qué preocupa a quien nos sigue, no para calcular ventas en otro barrio.»

Nora, que lo había oído todo, se quedó de mal humor, pero no tiró el informe. «Entonces el 38 % es una pista, no una cifra.» Pidió dos cosas: que el inventario dijera, fuente por fuente, qué servía y para qué, y que la encuesta propia contara como hogar a las personas que viven en la misma vivienda. Mientras tanto, Itxaso prepararía kits solo para los pedidos confirmados.

Preguntas. 1) ¿Qué afirmación del relato necesitaba una comprobación antes de convertirse en decisión? 2) ¿Qué información pidieron primero y por qué cambió la respuesta? 3) ¿Qué actuación provisional se acordó para aprender sin ocultar la incertidumbre? ¿Propondrías otra?

APRENDE CON LA IA

Unai pregunta a un asistente de IA si puede usar el padrón para saber cuántos hogares del barrio comprarían kits. Esta es la respuesta:

RESPUESTA DE LA IA

«Sí. El padrón es una fuente oficial y fiable, así que puedes usarlo sin más comprobaciones. Una fuente oficial responde automáticamente a cualquier pregunta comercial. Toma el número de empadronados y aplícale el porcentaje de compra que hayas encontrado.»

Tu tarea. Localiza el error, contrástalo con lo que has aprendido en los apartados 4.2 y 4.4 (cobertura, definición, suficiencia por objetivo) y escribe una respuesta corregida para Unai que diga para qué sirve el padrón y para qué no. Conserva la prueba que justifica tu corrección (la unidad de la ficha A y su definición).

Ojo con la IA. Un asistente puede confundir autoridad con pertinencia: pregúntale qué variable, población y periodo cubre la fuente. **Sin asistente:** trabaja con este recuadro.

ACTIVIDADES FINALES

Comprueba

12. Explica la diferencia entre una fuente fiable y una fuente pertinente, y aplícala a las fichas A y B del caso resuelto.
13. ¿Por qué el 38 % de la ficha B no se puede multiplicar por los hogares del barrio? Da dos motivos distintos.
14. Explica qué es una media móvil de tres semanas y qué componente de la serie de la Tabla 2 ayuda a ver y cuál esconde.
15. Explica qué son los metadatos y qué tres campos echarías en falta si recibieras la Tabla 1 sin ellos.

Aplica

16. Repite la práctica de Ruedalia cambiando una condición: el informe nacional es de 2019 y las 40 opiniones están en el perfil de la propia Ruedalia. Señala qué decisiones de aceptar, limitar o descartar cambian y cuáles se mantienen. Después, si tres empresas del polígono te dieran su número de ciclistas, ¿qué veredicto de la matriz cambiaría?
17. El tique medio de 2027 es de 14,30 €. Si el índice de kits de 2027 fuera 111,0, ¿cuál sería el tique en euros de 2023? ¿Ha subido o bajado en términos reales respecto a 2026?
18. Itxaso quiere saber si conviene preparar más kits los sábados. Diseña la búsqueda: qué fuente interna y qué fuente externa consultarías, con qué operadores, y qué registrarías de cada una.

Investiga

19. Localiza una fuente primaria o documentación técnica sobre archivo y recuperación de información (por ejemplo, una guía de un organismo público sobre metadatos, datos abiertos o gestión de documentos). Registra la fecha de consulta, el ámbito, la afirmación que te sirve para tu archivo y una limitación.

TU INVESTIGACIÓN AVANZA · PASO 4 · I04 · INVENTARIO Y LAGUNAS

Incorpora la entrega **I04 Inventario y lagunas** a tu dosier: para la investigación de tu plan I03, reúne al menos tres fuentes secundarias (una online oficial y una offline), con su registro de búsqueda; valóralas con las cuatro preguntas; pásalas a una tabla normalizada con metadatos; completa la matriz objetivo-evidencia-suficiencia y formula las preguntas pendientes con su técnica primaria.

Comprueba tu dosier: cada fuente tiene ficha, versión y fecha de consulta; los brutos están sin tocar; incluyes la evidencia (archivos y registro de cambios), la decisión, el responsable y la fecha de revisión.

☐ Hecho ☐ Revisado ☐ Guardado

OJO PROFESIONAL

Documento: nota de fuentes del estudio de kits (EK-2027/FS-v0.1)

Documento ficticio de La Despensa de Nora, S.L., redactado el miércoles 24/02/2027 para el inventario. Contiene **tres errores**.

N.º	Apartado	Contenido de la nota
1	Objetivo	Estimar la proporción de hogares del barrio que compraría kits (2027)
2	Fuente de demanda	Informe de una asociación de consumidores de otra ciudad (2022): el 38 % de los hogares compraría kits; se aplica al barrio
3	Mercado potencial	38 % de los 14.520 hogares del barrio = 5.518 hogares compradores
4	Clientela actual	Prueba de kits, 11/01-20/02/2027: 648 kits; 107 titulares compraron 435
5	Archivo	Padrón descargado el 22/02/2027; guardado como «padron_ultimo.xlsx», que se sobrescribe con cada actualización
6	Lagunas	Falta la intención de compra de los hogares del barrio: encuesta propia
7	Responsable	Equipo de análisis; revisa Adrián (cifras) y aprueba Nora
8	Próxima revisión	Viernes 26/02/2027, con el inventario completo

Encargo. Señala los tres errores, explica qué falla en cada uno con lo que has aprendido en la unidad y reescribe las líneas afectadas para que la nota se pueda utilizar. **Para pensar:** ¿qué dato añadirías a la línea 4 para que la clientela actual quede bien descrita?

N.º	Línea	Por qué es un error	Corrección
1			
2			
3			

AUTOEVALUACIÓN

1. Para el objetivo «hogares del barrio», el mejor registro de una fuente online es...
a) el enlace y nada más.
b) la ubicación, la fecha de consulta, el alcance y el dato extraído con su unidad.
c) una captura de la portada del organismo.

2. La ficha B (otra ciudad, 2022, voluntaria) para estimar la demanda del barrio se...
a) acepta, porque tiene 412 respuestas.
b) limita a hipótesis: otra población, otra fecha, otra definición y respuesta voluntaria.
c) acepta si se publica en una web oficial.

3. Una tabla normalizada sirve para...
a) comparar datos sin mezclar unidades ni definiciones, conservando el original.
b) sumar todas las cifras del inventario.
c) sustituir los datos brutos.

4. Diez informes sobre hogares de todo el país, para O2 del barrio, dan...
a) evidencia suficiente por su volumen.
b) evidencia parcial o ninguna, porque no cubren la población del objetivo.
c) un resultado cubierto si coinciden entre sí.

5. Al actualizar el padrón, lo correcto es...
a) sobrescribir el archivo anterior.
b) guardar la nueva versión con fecha, conservar la anterior y anotar el cambio.
c) borrar la versión de 2026 para no confundirse.

6. En el caso, la ficha C (435 kits de 107 titulares y 213 sin identificar) permite...
a) estimar cuántos hogares del barrio comprarían kits.
b) describir la clientela actual: 4,07 kits por titular identificado.
c) concluir que los 213 kits son de no clientes.

7. Txema dice: «28 de 40 opiniones: el 70 % quiere el servicio». Es incorrecto porque...
a) 28 de 40 es el 60 %.
b) las opiniones voluntarias de una red social no son una muestra de los trabajadores del polígono.
c) hacen falta 400 opiniones.

8. «Una fuente oficial responde automáticamente a cualquier pregunta comercial.» Es falso porque...
a) las fuentes oficiales no son fiables.
b) la autoridad no sustituye la pertinencia de la variable, el periodo y la población.
c) solo responden preguntas sobre empresas.

9. En el Ojo profesional, la línea 3 falla porque...
a) 14.520 son personas, no hogares, y el 38 % es de otra población.
b) el cálculo da 5.517.
c) debería usar el 38 % del índice de precios.

10. En tu entrega I04, ¿qué dato te obligaría a revisar la matriz de suficiencia?
a) Que una fuente use otra definición de la variable del objetivo.
b) Que una fuente tenga muchas páginas.
c) Que la fuente sea gratuita.

Ya sé...

- ☐ explicar las distinciones de los cinco apartados.
- ☐ auditar fuentes y registrar sus huecos.
- ☐ transferir el procedimiento a otro negocio (Ruedalia).
- ☐ detectar una conclusión no sustentada (autoridad no es pertinencia).
- ☐ entregar el inventario I04 como evidencia revisable.



UNIDAD 5

Investigación cualitativa

Bloque 2 · Fuentes e instrumentos

EN ESTA UNIDAD APRENDERÁS A…

- Elegir técnica (entrevista, grupo, observación o encuesta) según el dato.
- Preparar y conducir una entrevista en profundidad con preguntas abiertas, repreguntas y cierre neutral.
- Organizar una dinámica de grupo con moderación, turnos y registro, y reconocer la dominancia y la conformidad.
- Diseñar una pauta de observación que registre conductas y no motivos supuestos.
- Usar estímulos proyectivos y técnicas de creatividad para explorar significados, con sus límites.
- Codificar citas y observaciones con categorías explícitas y presupuestar la recogida en euros y en horas.

SITUACIÓN DE PARTIDA

Lunes 8 de marzo de 2027. Nora lee el acta del grupo de prueba que Unai moderó el viernes 05/03 en la zona de kits con seis clientes. La conclusión del acta es rotunda: «A los seis les encanta la idea de recoger el kit a las 19:00. El 100 % lo quiere».

Olvido, que tomaba notas en un rincón, no lo ve tan claro: «Unai empezó contando lo bien que nos iba a ir con la recogida de las siete y lo práctico que era. Desde ese momento nadie dijo otra cosa». En las tres entrevistas en profundidad de la semana anterior (E1, E2 y E3) nadie habló de las siete: hablaron de no saber cuándo recoger el kit y de horarios que no casan.

Nora quiere fijar ya el horario de recogida de los kits de hogar. Te pide revisar las dos fuentes y preparar el guion cualitativo con su codificación para el viernes 12/03.

PARA EMPEZAR

1. ¿Qué tiene que decidir Nora esta semana y qué tiene que comprobar antes de decidirlo?
2. ¿Qué sabemos de verdad por el grupo y por las tres entrevistas, y qué estamos suponiendo?
3. ¿Qué prueba permitiría revisar la primera respuesta del acta («el 100 % lo quiere»)?



Entrevista

5.1



Grupo

5.2



Observación

5.3



Proyectivas y creatividad

5.4



Registro y análisis

5.5

5.1 Entrevista en profundidad

La entrevista sirve para entender cómo vive una persona un problema. Si quien pregunta enseña su idea antes de escuchar, lo que oye es su propia idea devuelta.

Información primaria cualitativa. Ya sabes que la **información primaria** se obtiene para la investigación que se está haciendo, tanto en estudios comerciales como en estudios de opinión, y que hay técnicas cualitativas (entrevista, grupo, observación, proyectivas) y cuantitativas (encuesta). En grado superior el paso es otro: elegir la técnica por el **tipo de dato** que necesita la decisión y por el **sesgo** que cabe esperar de cada una. La **investigación cualitativa** busca razones, significados y palabras con pocos casos elegidos a propósito; no mide cuántos piensan algo. Sus resultados son **exploratorios**: abren hipótesis que después se miden con una encuesta.

Qué técnica para qué dato. La pregunta de Nora es de horarios: ¿a qué hora pueden recoger el kit los hogares? Detrás hay tres preguntas distintas:

Pregunta	Técnica más adecuada	Sesgo que hay que vigilar
¿Por qué no encaja el horario y cómo lo resolverían?	Entrevista en profundidad	Que la persona complazca a quien pregunta
¿Qué hacen los clientes ante la vitrina de kits?	Observación	Atribuir motivos a una conducta
¿Qué proporción de hogares recogería a cada hora?	Encuesta con muestra probabilística	Cobertura, no respuesta y redacción

Qué es. La **entrevista en profundidad** es una conversación individual, de 30 a 60 minutos, guiada por un **guion** de temas y no por un cuestionario cerrado. Quien entrevista escucha más que habla y pide ejemplos concretos. Sirve cuando el tema es personal o la presión de un grupo distorsionaría las respuestas.

El guion: de lo general a lo concreto. Un guion de entrevista no es una lista de preguntas para leer. Es una secuencia de temas con una **pregunta de apertura** amplia, preguntas abiertas para cada tema, **repreguntas** previstas y un **cierre**.

1. Presentación: quién hace la entrevista, para qué, cuánto dura, que no hay respuestas buenas ni malas, cómo se registra y para qué se usará lo que se diga.
2. Apertura: «Cuénteme cómo es una tarde normal entre semana, desde que sale de trabajar».
3. Temas: compras de comida, uso de comida preparada, momentos sin tiempo, cómo se entera de horarios y servicios.
4. Repreguntas: «¿Me pone un ejemplo de la última vez?», «¿Qué pasó entonces?», «¿Qué quiere decir con "tarde"?».
5. Concepto, solo al final: se describe el kit sin adjetivos y se pide la reacción.

1. Cierre: «¿Hay algo que no le haya preguntado y que debería saber?»; agradecimiento.

Preguntar sin sugerir. «¿Cómo organiza la cena los días de trabajo?» no orienta la respuesta; «¿Verdad que le vendría bien una cena ya hecha?», sí. La repregunta usa las palabras de la persona.

CLAVE · ESCUCHAR ANTES DE PRESENTAR

En una entrevista, la oferta se presenta al final y sin adjetivos. Antes de escuchar, cualquier entusiasmo de quien pregunta se convierte en respuesta: la persona tiende a decir lo que cree que el otro quiere oír (**sesgo de deseabilidad social** o de complacencia).

Las tres entrevistas del caso. El equipo hizo tres entrevistas los días 2, 3 y 4 de marzo, en la zona de kits fuera del horario de preparación o en la oficina de la persona: E1 (madre con dos hijos, 41 años, vive a 300 m), E2 (administrativo de una oficina del polígono, 33 años) y E3 (enfermera a turnos, 52 años). Se eligieron a propósito por ser tres situaciones distintas, no como muestra del barrio; las tres aceptaron por escrito la grabación del audio.

LA NORMA DICE

Reglamento (UE) 2016/679 (RGPD), de 27 de abril de 2016, art. 5.1 c): los datos personales serán «adecuados, pertinentes y limitados a lo necesario en relación con los fines para los que son tratados («minimización de datos»)». En la transcripción, un código (E1) y un perfil, no el nombre ni el teléfono. Consulta la versión vigente.

ACTIVIDADES

1. Compara entrevista, observación y encuesta frente a la pregunta de horarios de Nora. Para cada técnica indica qué dato obtiene, qué sesgo cabe esperar y en qué momento del estudio la usarías.
2. Reescribe sin sugerir respuesta estas dos preguntas de un guion: a) «¿No le parece que recoger a las siete es lo más cómodo?»; b) «¿Qué es lo que más le gusta de nuestros kits?». Añade una repregunta para cada una.

5.2 Dinámica de grupo

Un grupo produce ideas que una persona sola no tendría. También produce acuerdos que nadie tenía antes de entrar.

Qué es. La **dinámica de grupo** (o **grupo de discusión**, *focus group*) reúne a varias personas con un perfil acordado para hablar de un tema guiadas por una persona que **modera**. Ya conoces el formato habitual: un grupo pequeño, de unos 6 a 10 participantes, con una moderadora o un moderador. Sirve para explorar vocabulario, reacciones a una idea, ventajas e inconvenientes que se rebaten entre sí y propuestas que nacen de la conversación. No sirve para saber cuántas personas piensan algo.

Organizar el grupo. Antes de la sesión se fijan por escrito:

- **Perfil de participantes** y cómo se captan (por ejemplo, hogares con hijos que compran comida preparada al menos una vez por semana). Se evita que se conozcan entre sí o que trabajen para la tienda.
- **Pauta de moderación**: temas en orden, tiempo de cada uno, estímulos que se enseñarán y en qué momento.
- **Papeles**: quien modera no toma notas; una segunda persona registra turnos, citas literales y lenguaje no verbal.
- **Registro**: grabación con permiso informado, hoja de turnos y acta con citas codificadas.

Moderar. Se abre con una pregunta amplia, se reparte la palabra y se contrasta («¿Alguien lo vive de otra manera?»), sin opinar ni elogiar la oferta, que llega al final.

Dos riesgos que se registran. La **dominancia** aparece cuando una o dos personas acaparan la conversación y las demás se callan o asienten. La **conformidad** aparece cuando las personas ajustan su opinión a la del grupo o a la que perciben en quien modera. Las dos se pueden ver en la hoja de turnos.

Registro 1 · Hoja de turnos del grupo del viernes 05/03/2027

(ficticio; 31 intervenciones en 75 minutos)

Participante	Perfil	Intervenciones	% del total
P1	hogar con hijos	3	9,7
P2	hogar con hijos	14	45,2
P3	pareja que trabaja hasta tarde	4	12,9
P4	persona que vive sola	5	16,1
P5	hogar con persona mayor	2	6,5
P6	pareja que trabaja hasta tarde	3	9,7

Los porcentajes no suman 100 por el redondeo. P2 tuvo casi la mitad de las intervenciones (14 de 31) y P5 solo dos. P2 opinó sobre las 19:00 justo después de que Unai la presentara como «lo más práctico», y P5 dijo al salir: «A mí a las siete no me viene, pero no quise llevar la contraria».

Si se repite el grupo. Ledicia, una moderadora con experiencia que colabora con el equipo, presentará el horario como una de cuatro opciones, sin adjetivos y después de hablar de las rutinas de la tarde; antes de hablar, cada participante anotará su preferencia en una tarjeta.

Consenso no es prevalencia. Que seis personas digan lo mismo en un grupo no significa que el 100 % de los hogares piense así. Primero, porque seis personas elegidas a propósito no representan al barrio. Segundo, porque el acuerdo de un grupo puede ser efecto de la dinámica (dominancia, conformidad, una idea lanzada por quien modera). Lo que sí aporta el grupo son argumentos, palabras y dudas que después se convierten en preguntas medibles.

ACTIVIDADES

3. Con la hoja de turnos, calcula qué parte de las intervenciones acumulan P2 y P4 juntas y explica qué señal de dominancia ves. Propón dos técnicas de moderación para equilibrar los turnos en la siguiente sesión.
4. Redacta la pauta de una segunda dinámica de grupo sobre el horario de recogida (perfil, captación, temas con tiempos, momento del estímulo y registro). Incluye una alternativa a la dinámica y la condición que te haría elegir esa alternativa.
5. Reescribe la conclusión del acta de Unai («El 100 % lo quiere») para que diga solo lo que el grupo permite afirmar.

5.3 Observación

Lo que se ve es una conducta. El motivo de esa conducta no se ve: se pregunta o se deja como pregunta abierta.

Qué es. La **observación** registra conductas tal como ocurren, sin preguntar. Puede ser **abierta** o **encubierta** (la persona sabe o no que la observan), **participante** o **no participante** (quien observa forma parte o no de la situación), y **estructurada** (con una pauta cerrada) o **no estructurada** (notas de campo libres). En una tienda la más útil suele ser la estructurada y no participante: alguien en un lateral del mostrador anota qué hacen los clientes ante un expositor.

Conducta, no motivo. Una **conducta observable** es algo que dos observadores verían igual: «se detiene ante la vitrina más de cinco segundos», «coge un kit y lee la etiqueta», «pregunta en el mostrador», «se va sin comprar». «Le pareció caro», «no le gustó» o «tenía prisa» son interpretaciones: pueden ser ciertas o no, y la observación sola no permite saberlo. Si el motivo importa, se combina con una pregunta breve al salir o con una entrevista.



La pauta antes de observar. Adrián ya contaba barras de pan con palotes en una libreta; ahora la pauta tiene que decir de antemano **qué** se observa, **dónde**, **cuándo** y **cómo se anota**:

Elemento de la pauta	En la vitrina de kits de La Despensa
Objetivo	Ver qué hacen los clientes ante los kits de la tarde
Lugar y posición	Lateral del mostrador, sin tapar la vitrina

Elemento de la pauta	En la vitrina de kits de La Despensa
Periodo	Martes 09/03/2027, de 17:00 a 19:00
Unidad	Persona que se detiene ante la vitrina
Conductas	Se detiene · coge un kit · pregunta · compra · se va sin comprar
Registro	Palotes por conducta y nota literal de lo que se pregunta

Registro 2 · Observación en la vitrina de kits

(ficticio; martes 09/03/2027, de 17:00 a 19:00)

Conducta observada	Personas
Se detiene ante la vitrina (5 s o más)	11
Coge un kit y lee la etiqueta	6
Pregunta en el mostrador por los kits	4
Compra un kit	3
Se va sin comprar tras detenerse	8

Comprobación: de las 11 personas que se detienen, 3 compran y 8 se van sin comprar (3 + 8 = 11). De las 4 preguntas, Adrián anota literalmente que 3 fueron sobre la hora de recogida o de reparto y 1 sobre alérgenos. Lo que el registro **permite** decir: en esa franja, de las 11 personas que se detuvieron, 8 se fueron sin comprar, y la pregunta más repetida fue la del horario. Lo que **no permite** decir: por qué se fueron las 8 ni qué pasa otros días u otras horas.



ACTIVIDADES

6. Clasifica como conducta observable o como interpretación y reescribe las interpretaciones: a) «mira el precio»; b) «le parece caro»; c) «se va sin comprar»; d) «no le convence el kit»; e) «pregunta a qué hora se recoge».

7. Diseña la pauta de una segunda observación en la vitrina de kits un sábado por la mañana: objetivo, posición, periodo, unidad, cinco conductas con su definición y forma de registro. Di qué técnica añadirías si Nora quiere conocer el motivo de quienes se van sin comprar.

5.4 Técnicas proyectivas y creatividad

A veces la gente no sabe o no quiere decir por qué hace algo. Un estímulo indirecto ayuda a que aparezcan significados que la pregunta directa no saca.

Técnicas proyectivas. Una **técnica proyectiva** presenta un estímulo ambiguo y pide a la persona que lo complete o lo interprete; se supone que, al hacerlo, proyecta sus propias actitudes. Las más usadas en estudios comerciales:

Técnica	Cómo se aplica al kit	Qué puede aportar
Completar frases	«Una persona que compra la cena hecha es…»	Prejuicios sobre la comida preparada
Tercera persona	«¿Qué cree que piensan sus vecinos de estos kits?»	Opiniones que no se dicen en primera persona
Asociación de palabras	Kit, recoger, siete de la tarde: primera palabra que venga	Vocabulario y connotaciones
Viñeta o bocadillo vacío	Dibujo de una persona ante la tienda cerrada; ¿qué piensa?	Situaciones de frustración
Collage o tarjetas	Elegir imágenes que representen «una tarde entre semana»	Emociones y rutinas

Límites de interpretación. Lo que sale de una proyectiva se interpreta y dos analistas pueden leerlo distinto. Se registra literalmente y se contrasta: si en la viñeta de la tienda cerrada aparece «otra vez llego tarde», la hipótesis del horario se refuerza, pero no se prueba. Además, una proyectiva se aplica con cuidado: el estímulo no puede ridiculizar a nadie ni provocar incomodidad, y la persona puede no contestar.

Técnicas de creatividad. Generan ideas, no miden. La **tormenta de ideas** pide cantidad sin juzgar y después agrupa y filtra; **seis sombreros** reparte puntos de vista; **SCAMPER** recorre la oferta con verbos (sustituir, combinar, adaptar, eliminar…). Una tormenta de ideas con Itxaso y Olvido propuso taquilla refrigerada, aviso por mensaje cuando el kit está listo y reparto desde las 20:00: cada idea es una **hipótesis de solución** que hay que comprobar con clientes y con la capacidad de la zona de kits.

CLAVE · UN ESTÍMULO, UNA INTERPRETACIÓN CON SU EVIDENCIA

En el registro, la frase completada o el collage van en una columna; lo que el equipo cree que significan, en otra; y la evidencia que lo apoya (otra cita, una observación), en una tercera. Si no hay evidencia, la interpretación se queda como hipótesis.

EN TU SECTOR

Marketing y publicidad: antes de una campaña se prueban conceptos y piezas con grupos y proyectivas para ver qué significa un mensaje, no cuántos lo recuerdan. **Gestión de ventas y espacios comerciales:** la observación estructurada en tienda (recorridos, paradas ante un lineal, tiempo en caja) orienta la colocación del producto y los turnos del personal.

ACTIVIDADES

8. Prepara para el caso de los kits un guion de entrevista (apertura, tres temas con su repregunta, presentación del concepto y cierre), una pauta de grupo (perfil, temas y tiempos), una ficha de observación (cinco conductas definidas) y un estímulo proyectivo o de creatividad. Simula al menos una de las técnicas con un compañero o una compañera y separa en tu registro observación, cita e interpretación.

9. Elige una de las tres ideas de la tormenta de ideas (taquilla, aviso por mensaje, reparto desde las 20:00) y escribe qué tendrías que comprobar antes de proponerla a Nora: una prueba con clientes y un dato de capacidad que pedirías a Itxaso.

5.5 Guion, registro y análisis

El análisis cualitativo se puede revisar si cada paso queda escrito: qué se dijo, qué código se le dio, qué se interpreta y con qué evidencia.

Codificar. Tras transcribir lo necesario, cada cita lleva su fuente (E1, P3, observación del 09/03) y su fecha. **Codificar** es asignar a cada cita u observación una o varias **categorías** (códigos) definidas por escrito en un **libro de códigos**. Cada código tiene nombre, definición con su criterio de inclusión y un ejemplo. Pueden salir del objetivo (deductivos) o de las transcripciones (inductivos).

Libro de códigos del estudio de kits (versión del 12/03/2027)

Código	Definición y criterio de inclusión	Ejemplo
INF	Barrera de información: no conoce horario, forma o lugar de recogida	«No sé cuándo…»
HOR	Compatibilidad horaria: sus horarios no casan con los de la tienda	«Salgo tarde»
INT	Interés declarado por el kit o por probarlo	«Me interesa»

La tabla cita · código · interpretación · evidencia. Es el recurso central del análisis: separa lo que se dijo de lo que el equipo cree que significa.

Cita (fuente)	Código	Interpretación	Evidencia
«No conozco el horario de recogida.» (E1)	INF	No conoce el horario	E2 dice algo parecido
«Cuando llego, la tienda ya está cerrada.» (E3)	HOR	La franja de tarde no le sirve	E2 tiene otro choque horario

Saturación y alcance. En un estudio cualitativo se sigue entrevistando hasta que las nuevas entrevistas ya no aportan códigos nuevos (**saturación**). Con tres entrevistas no se ha alcanzado. Contar en cuántas entrevistas aparece un código ordena el análisis, pero no es un porcentaje del barrio.

Calidad, coste y utilidad. Elegir técnica también es presupuestar. Una técnica barata en euros puede ser cara en horas, y una barata en las dos puede no responder al objetivo. Para la siguiente fase, el equipo compara dos alternativas de recogida de datos para un objetivo: entender por qué el horario no encaja en cada tipo de hogar. Agradecimiento: un K02 por participante (coste de referencia 5,80 €).

Cuadro 1 · Dos alternativas de recogida de datos

(euros sin IVA; horas del equipo; B: 2 h de captación + 2 personas × 1,5 h; merienda, 2 K02)

Concepto	A · 6 entrevistas	B · 1 grupo de 8
Agradecimientos (€)	6 × 5,80 = 34,80	8 × 5,80 = 46,40
Merienda de la sesión (€)	—	2 × 5,80 = 11,60
Impresión de guiones y hojas (€)	6,00	4,00
Total en euros	40,80	62,00
Preparación (h)	3	4
Captación y recogida de datos (h)	6 × 1 = 6	2 + 2 × 1,5 = 5
Transcripción y análisis (h)	6 × 1,5 = 9	8
Total en horas	18	17

A cuesta 21,20 € menos en euros y una hora más de trabajo del equipo; ninguna de las dos incluye el coste de la encuesta posterior, que es la que medirá cuántos hogares tienen cada barrera. Pero la decisión no se toma por el coste: el objetivo es entender barreras individuales de horario, un tema que depende de la rutina de cada hogar y en el que el grupo del día 5 ya mostró conformidad. La alternativa A responde mejor; el grupo B sería preferible si el objetivo fuera generar ideas de servicio o probar nombres y mensajes.

Dato, interpretación y recomendación. El informe los separa: **dato** («las tres personas hablan del horario»), **interpretación** («la tarde no sirve a quien sale tarde») y **recomendación** («medir la franja preferida antes de fijar la recogida»).

PARA SABER MÁS

El Código Internacional ICC/ESOMAR de investigación de mercados, social y de opinión y de análisis de datos (edición de 2025) no es una ley: es el código de autorregulación de la profesión. Pide, entre otras cosas, investigar con honestidad y transparencia, tratar a los participantes con respeto y sin perjudicarles, explicarles con claridad cómo se usan sus datos y no confundir investigar con vender. Consulta la versión vigente.

ACTIVIDADES

10. Presupuesta dos alternativas de recogida de datos para este objetivo: «saber qué hacen los clientes ante la vitrina de kits a última hora de la tarde». Alternativa 1: observación estructurada durante cuatro tardes de 2 h (pauta, 2 h; análisis, 3 h; impresión de fichas, 2,00 €). Alternativa 2: cuatro entrevistas telefónicas de 45 minutos (guion, 2 h; análisis, 1 h por entrevista; agradecimiento, un K02 por persona, recogido en tienda). Calcula euros y horas y argumenta cuál responde mejor.

11. Define un código nuevo, PRE (precio), con su criterio de inclusión y un ejemplo, y di si alguna cita del caso resuelto lo llevaría.

12. Con el Cuadro 1, calcula cuánto costaría la alternativa A si hicieras ocho entrevistas en lugar de seis (mismos costes unitarios y preparación; impresión, 1,00 € por entrevista). ¿Cambiaría tu elección? ¿Qué señal indicaría saturación y que sobran entrevistas?

CASO RESUELTO · PASO A PASO

Tres entrevistas sobre la recogida de kits

Caso ficticio. Viernes 12 de marzo de 2027. El equipo codifica las entrevistas E1, E2 y E3 para entender por qué no encaja el horario de recogida; Adrián lo comprueba.

Paso 1. Decisión, datos y alcance. Decisión: horario de recogida de los kits de hogar. Datos: seis citas de tres entrevistas elegidas a propósito. Alcance: explorar barreras y formular hipótesis; no estimar cuántos hogares las tienen.

Paso 2. Codificación con el libro de códigos

Cita (fuente)	Código	Interpretación
«No conozco el horario de recogida.» (E1)	INF	Desconoce el horario (¿dónde buscó?)
«Si fuera a la salida del colegio, lo probaría; ahora no llego.» (E1)	INT, HOR	Interés condicionado a una franja
«No sé cuándo puedo pasar a recoger el kit.» (E2)	INF	Duda sobre el procedimiento
«En la oficina comemos a las dos; a esa hora no puedo salir.» (E2)	HOR	Su hora libre no coincide
«Me interesa, pero salgo tarde.» (E3)	INT, HOR	Interés frenado por el horario
«Cuando llego, la tienda ya está cerrada.» (E3)	HOR	La tarde no le sirve

Recuento por entrevista: INF en 2 de 3 (E1, E2); HOR en 3 de 3; INT en 2 de 3 (E1, E3). Adrián codificó las seis citas sin ver la tabla y coincidió en 5 de 6 (83,3 %); la diferencia (cita 2 de E1, solo INT para él) se resolvió con la definición de HOR.

Paso 3. Comprobación e interpretación. Permite concluir que en las tres situaciones aparece una barrera horaria y en dos una de información, y que el interés está condicionado. No permite decir que «el 67 % del barrio no conoce el horario» ni que «a todos les falla el horario»: tres testimonios elegidos a propósito no estiman ningún porcentaje poblacional.

RESULTADO

Citas, códigos e interpretación quedan separados y cada interpretación tiene su evidencia. Hipótesis para medir en la encuesta: el horario de recogida es una barrera y la información sobre él no llega. Recomendación: no fijar aún las 19:00; publicar ya el horario actual en tienda y web (barato y reversible) y medir en la encuesta la franja preferida.

Comprueba que lo has entendido. a) ¿Por qué se cuenta por entrevista y no por cita? b) ¿Y si una cuarta entrevista no aporta códigos nuevos? c) ¿Qué frase del resultado es dato, interpretación o recomendación?

PRÁCTICA

Observación en el mostrador de Ruedalia

Txema quiere saber por qué algunos clientes particulares entran en Ruedalia y se van sin pedir nada. Ousmane observa desde el banco de trabajo un martes, de 9:00 a 10:00, con una ficha en la que anota lo que ve. Su registro dice:

Hora	Lo que se observa
9:12	Hay una cola de cuatro personas en el mostrador. Una persona entra, mira la cola y sale
9:20	Con la misma cola, otra persona entra, mira el reloj y sale
9:41	Una persona pregunta el precio de la revisión básica, escucha la respuesta y sale sin pedir cita

En la nota final Ousmane ha escrito: «Se van porque somos caros y lentos».

Encargo. 1) Separa en el registro las conductas observables de las interpretaciones. 2) Codifica cada conducta con categorías definidas por ti (por ejemplo, COLA, PRECIO). 3) Escribe qué permite concluir la observación y qué no. 4) Propón una forma de conocer el motivo sin inventarlo y una decisión provisional para Txema.

Plantilla IC-P05 · Observación en Ruedalia

Campo	Tu respuesta
Datos y fuentes con unidad y periodo	
Procedimiento o cálculo	
Resultado y evidencia	
Decisión y limitaciones	

Lista de cotejo

☐ Cada conducta está descrita sin motivo.
☐ Los códigos tienen definición.
☐ Las conclusiones citan su evidencia.
☐ La decisión dice qué la haría cambiar.

LECTURA Y DEBATE

La primera vez que hablé demasiado



Relato ficticio. Ledicia lleva quince años moderando grupos de discusión. El lunes 8 de marzo pasa por La Despensa para preparar la segunda sesión y Unai le enseña su acta. Ledicia sonríe: «Yo hice lo mismo en mi primer grupo».

Era un estudio para una cadena de cafeterías que quería lanzar un bocadillo vegetal. «Me habían contado la idea con tanta ilusión que la presenté a los diez minutos: "Es sano, es rápido y está riquísimo". A partir de ahí, el grupo fue un coro». El informe dijo que el producto gustaba mucho; la empresa lo sacó a la venta y, al mes, casi no se vendía.

Lo que obligó a revisarlo no fue una opinión, sino el vídeo de la sesión. Su supervisora le pidió que contara cuántas veces había hablado ella antes de la primera opinión del grupo sobre el bocadillo: siete. Y que mirara las caras: dos personas habían cruzado los brazos al oír «vegetal», pero no dijeron nada. «Mi decisión fue presentar la idea pronto para no perder tiempo. La comprobación fue mirar mi propia grabación».

Desde entonces, Ledicia tiene tres reglas: el concepto llega en el último tercio de la sesión; se presenta sin adjetivos y junto a otras opciones; y cada participante escribe su opinión en una tarjeta antes de decirla en voz alta. «Y el acta dice siempre quién habló primero», añade, mirando la de Unai.

Preguntas. 1) ¿Qué afirmación del acta de Unai necesita una comprobación antes de convertirse en decisión? 2) ¿Qué información pedirías primero (la grabación, la hoja de turnos, las tarjetas…) y por qué cambiaría la respuesta? 3) Propón una actuación provisional para Nora que le permita aprender sobre el horario sin ocultar la incertidumbre.

APRENDE CON LA IA

Unai pega en un asistente de IA las seis citas del caso resuelto y le pide «un titular para el informe». La respuesta es esta:

> **RESPUESTA DE LA IA**
> «Dos de los tres entrevistados no saben cuándo recoger el kit: el 66,67 % del barrio tiene ese problema.»

Tu tarea. Localiza el error, contrástalo con el procedimiento de la unidad y escribe una respuesta corregida con su límite. Conserva la prueba que justifica la corrección (la tabla de codificación, el recuento por entrevista o la forma de selección de las personas entrevistadas).

Ojo con la IA. Un porcentaje con dos decimales parece preciso, pero la precisión del número no dice nada de cómo se eligieron las personas.

ACTIVIDADES FINALES

Comprueba

13. Resume la distinción entre conducta observable e interpretación, y la distinción entre consenso de un grupo y prevalencia en la población. Aplica las dos al caso resuelto.
14. Explica con un ejemplo del caso por qué el moderador no debe presentar la oferta al principio de un grupo y qué condición te haría aceptar el resultado de un grupo como pista útil.
15. ¿En qué se diferencian el libro de códigos y la tabla cita · código · interpretación · evidencia?

Aplica

16. Repite la práctica de Ruedalia cambiando una condición: la observación se hace un jueves de 16:00 a 17:00, sin cola, y dos personas preguntan el precio y se van. Señala qué resultado cambia (qué código aparece y cuál desaparece) y qué se mantiene (lo que la observación no permite saber).
17. Con el Registro 2 de la observación de la vitrina, redacta tres frases para Nora: un dato, una interpretación con su evidencia y una recomendación con su condición de revisión.
18. Codifica con INF, HOR e INT estas dos citas de una cuarta entrevista (E4, repartidor de 29 años) y di si hay saturación: «Yo los domingos cocino para toda la semana» y «Si me avisan al móvil cuando está listo, paso».

Investiga

19. Localiza una fuente primaria o una documentación técnica sobre guion, registro y análisis cualitativo (un código profesional de la investigación de mercados, la guía de un organismo público de estadística o de un grupo universitario sobre entrevistas o codificación). Registra fecha de consulta, ámbito, una afirmación que sustenta lo que has aprendido en la unidad y una limitación de la fuente.
20. Busca en la web de la Agencia Española de Protección de Datos qué información hay que dar antes de grabar una entrevista. Anota la fecha de consulta.

TU INVESTIGACIÓN AVANZA · PASO 5 · I05 · GUION CUALITATIVO Y CODIFICACIÓN

Para la investigación de tu negocio, prepara un guion de entrevista en profundidad y una pauta de observación (o de grupo) ligados a tu pregunta de decisión. Haz al menos dos entrevistas o una observación de prueba, transcribe las citas pertinentes, escribe tu libro de códigos y construye la tabla cita · código · interpretación · evidencia. Incluye la decisión que propones, quién es responsable de cada técnica y la fecha en que la revisarás. Entrega los archivos de registro, no solo una presentación.
Comprueba tu entrega: ninguna conclusión da porcentajes de la población; cada interpretación tiene su evidencia; las personas aparecen con código y sin datos innecesarios.
☐ Hecho ☐ Revisado ☐ Guardado

OJO PROFESIONAL

Documento: informe de campo cualitativo, versión 1

Documento ficticio de La Despensa de Nora, S.L., borrador preparado por Unai el viernes 12 de marzo de 2027 para Nora. Contiene **cinco errores**.

N.º	Apartado	Texto del borrador
1	Técnicas	Tres entrevistas en profundidad (E1, E2, E3), un grupo de seis personas (05/03) y una observación en la vitrina de kits (09/03, de 17:00 a 19:00)
2	Moderación	El moderador abrió el grupo explicando que la recogida a las 19:00 «es lo más práctico y va a encantar» para animar la conversación
3	Observación	De las 11 personas que se detuvieron ante la vitrina, 8 se fueron sin comprar porque los kits son caros
4	Registro	Se adjuntan la hoja de turnos del grupo y la tabla cita · código · interpretación · evidencia
5	Resultados	INF aparece en 2 de 3 entrevistas: el 66,67 % de los hogares del barrio no conoce el horario de recogida
6	Datos de las personas	Transcripción de E1 con su nombre, su dirección y su teléfono para volver a llamarla
7	Conclusión	El grupo confirma que el 100 % de los clientes quiere recoger a las 19:00: se fija ese horario desde el lunes 15/03

Encargo. Señala los cinco errores, explica por qué cada uno es un error con lo que has aprendido en la unidad y reescribe el informe para que Nora lo pueda utilizar. Di qué dos líneas están bien.

N.º	Línea	Por qué es un error	Corrección
1			
2			
3			
4			
5			

AUTOEVALUACIÓN

1. Nora quiere saber qué proporción de hogares recogería el kit a cada hora. ¿Qué técnica responde?

a) Tres entrevistas bien elegidas.
b) Una encuesta con muestra probabilística.
c) Dos horas de observación.

2. En un grupo, P2 tiene 14 de 31 intervenciones y todos acaban apoyando la idea del moderador.

a) Hay dominancia y posible conformidad: no mide cuántos lo piensan.
b) El 100 % de los hogares la quiere.
c) El consenso valida el grupo.

3. ¿Cómo se anota «se fue porque la revisión es cara» en una pauta?

a) Igual.
b) «Pregunta el precio y sale sin pedir cita.»
c) «Cliente insatisfecho con los precios.»

4. ¿Qué prepara de forma completa la fase cualitativa?

a) Un cuestionario cerrado.
b) Guion de entrevista, pauta de grupo, ficha de observación y un estímulo proyectivo o creativo.
c) Un grupo y su acta sin turnos.

5. Entrevistas: 40,80 € y 18 h; grupo: 62,00 € y 17 h; objetivo: barreras horarias individuales.

a) El grupo, por las horas.
b) Las entrevistas, por el objetivo.
c) La más barata en euros.

6. HOR aparece en 3 de 3 entrevistas e INF en 2 de 3.

a) Hay barrera horaria en las tres situaciones, sin porcentaje del barrio.
b) El 66,67 % del barrio no conoce el horario.
c) El 100 % tiene problemas de horario.

7. En Ruedalia, con cola de cuatro personas, dos entran y salen, y una sale tras preguntar el precio.

a) «Somos lentos y caros.»
b) Dos salidas con cola y una tras preguntar el precio, sin motivo.
c) «Tres clientes perdidos por el precio.»

8. «Dos de tres entrevistados no saben cuándo recoger: el 66,67 % del barrio tiene ese problema.» ¿El error?

a) Es el 66,66 %.
b) Tres entrevistas elegidas a propósito no estiman un porcentaje poblacional.
c) Debería decir 67 %.

9. El moderador abrió diciendo que la recogida a las 19:00 «va a encantar».

a) Concepto al final, sin adjetivos y entre otras opciones.
b) Repetir el elogio al final.
c) Quitar el párrafo del informe.

10. Tu I05 propone publicar el horario en la web. ¿Qué dato te haría revisarlo?

a) Ninguno.
b) Que la encuesta muestre que quien no compra ya conoce el horario y su problema es la franja.
c) Que a Unai le guste otra idea.

Ya sé...

- ☐ explicar las diferencias entre entrevista, grupo, observación, proyectivas y encuesta y elegir según el dato.
- ☐ ejecutar la codificación de citas con libro de códigos y tabla cita · código · interpretación · evidencia.
- ☐ transferir el procedimiento de observación a otro negocio, como Ruedalia.
- ☐ detectar una conclusión no sustentada, como un porcentaje del barrio a partir de tres entrevistas.
- ☐ entregar una evidencia revisable: guion, registro, códigos y decisión con fecha de revisión.



UNIDAD 6

Encuestas y cuestionarios

Bloque 2 · Fuentes e instrumentos

EN ESTA UNIDAD APRENDERÁS A…

- Comparar la encuesta personal, postal, telefónica y en línea por cobertura, coste, control y sesgos.
- Distinguir CAPI, MCAPI, CATI y CAWI por dispositivo, entrevistador y forma de captura.
- Redactar preguntas neutrales con una idea, un periodo y opciones equilibradas, e informar a quien responde.
- Programar filtros y saltos y hacer un pretest documentado con duración, rutas y registro de cambios.
- Elegir la técnica según el objetivo, el plazo y el presupuesto, y caracterizar los paneles.

SITUACIÓN DE PARTIDA

Lunes 15 de marzo de 2027. El equipo revisa con Nora y Unai el cuestionario v0.1 del estudio de kits. Unai lo tiene claro: «Lo subimos a un formulario en línea, mandamos el enlace a los clientes con correo y en una semana lo tenemos. Coste cero». Olvido, que ha bajado del obrador con una bandeja, no lo ve tan claro: «Media clientela de la tarde no ha abierto un correo en su vida. ¿A esa gente no le preguntamos?».

Mientras, Adrián ha probado el formulario en su móvil como si no hubiera comprado nunca un kit, y le ha salido la pregunta «¿No cree que nuestro excelente kit es barato y saludable?». Nora quiere tener los datos en cuatro semanas de abril y te pide que elijas la modalidad, corrijas el cuestionario y lo pruebes antes de lanzarlo.

PARA EMPEZAR

1. ¿Qué tiene que decidir Nora y qué hay que comprobar antes de lanzar la encuesta?
2. ¿Qué sabemos de quién tiene correo y quién no, y qué estamos suponiendo al decir «coste cero»?
3. ¿Qué prueba permitiría revisar la primera respuesta de Unai antes de gastar las cuatro semanas?



Modalidades

6.1



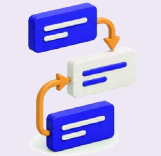
CAPI, MCAPI, CATI, CAWI

6.2



Redacción y escalas

6.3



Saltos y pretest

6.4



Paneles y elección

6.5

6.1 Modalidades de encuesta

La modalidad decide a quién se llega antes de que nadie conteste. Lo barato por contacto puede salir caro en sesgo.

La encuesta ad hoc. Una **encuesta** recoge datos de muchas personas con un **cuestionario** igual para todas, para poder contar y comparar. Es **ad hoc** cuando se diseña para un estudio concreto, como el de kits. Ya sabes que puede ser personal, telefónica o en línea; aquí se comparan las cuatro modalidades clásicas con criterios de grado superior: **cobertura** (a quién puede llegar), **coste** por respuesta, **control** de quién contesta y cómo, y **sesgos** previsibles.

Matriz de las cuatro modalidades

Modalidad	Cobertura y control	Coste y plazo	Sesgos frecuentes
Personal	Alta si hay acceso al lugar; control alto	Alto coste; plazo medio	Efecto del entrevistador
Postal (por correo)	Quien tiene dirección registrada; control bajo	Bajo coste; plazo largo	Baja respuesta; autoselección
Telefónica	Quien tiene teléfono y contesta; control medio	Coste medio; rápida	Rechazo; preguntas cortas
En línea	Quien usa internet y da su correo; control bajo	Muy bajo coste; rápida	Cobertura y autoselección

Cobertura: a quién llega cada modalidad. El **marco** de la encuesta son los 120 hogares que han comprado algún kit y figuran en el registro de clientes de Itxaso a 12/03/2027. No todos se pueden alcanzar igual:

Registro 1 · Contacto de los 120 hogares compradores de kits

(ficticio; a 12/03/2027)

Dato de contacto	Hogares	% de 120	De 65 años o más
Correo autorizado (estudio)	78	65,0	6
Solo teléfono	34	28,3	22
Sin contacto (en tienda)	8	6,7	8
Total	120	100,0	36

Ejemplo. La cobertura de cada modalidad

- En línea (correo): 78 de 120 = **65,0 %**. Quedan fuera 42 hogares, y 30 de ellos tienen 65 años o más.
- Telefónica: los que tienen correo también dieron teléfono, así que 78 + 34 = 112 de 120 = **93,3 %**.
- Personal en tienda: Itxaso calcula que 96 hogares recogen en tienda al menos una vez al mes: **80,0 %**.
- Postal: 64 hogares tienen dirección de reparto registrada: **53,3 %**.

Comprobación del perfil: en el registro hay 36 hogares de 65 años o más (30,0 %); entre los que tienen correo solo hay 6 de 78 (7,7 %). Una encuesta solo en línea no es solo «más pequeña»: está inclinada hacia los hogares más jóvenes. Por eso, antes de elegir, se compara el perfil del marco con el de cada modalidad: edad, canal de compra y franja de recogida.

CLAVE · COBERTURA NO ES RESPUESTA

La **cobertura** dice a quién puede llegar la modalidad; la **tasa de respuesta**, cuántos de los contactados contestan. Una encuesta en línea con muchas respuestas puede tener mala cobertura: Internet no llega a todos los clientes por igual.

ACTIVIDADES

1. Construye la matriz de cobertura, coste, control y sesgos de las cuatro modalidades para el estudio de kits, con un dato del Registro 1 en cada fila. ¿Qué modalidad no usarías sola y por qué?
2. Calcula la cobertura de una encuesta mixta: en línea para los 78 hogares con correo y personal en tienda para los que no tienen correo, sabiendo que 36 de esos 42 recogen en tienda. ¿Qué hogares siguen fuera?

6.2 CAPI, MCAPI, CATI y CAWI

Un ordenador no cambia el canal, pero cambia cómo se captura y se valida cada respuesta.

Asistencia informática y canal. Las siglas combinan el **canal** (cara a cara, teléfono o web) con la **asistencia por ordenador** (*computer assisted*). En la versión en papel (PAPI, *paper and pencil*), la persona que encuesta lee, anota a mano y después alguien graba los datos; en las asistidas, el programa muestra la pregunta que toca, aplica los saltos y comprueba la respuesta en el momento.

Sigla	Canal y dispositivo	¿Hay entrevistador?	Captura y validación
CAPI	Cara a cara, con ordenador o tableta	Sí, presente	Registra en el acto; saltos y rangos automáticos
MCAPI	Cara a cara, con móvil o tableta	Sí, presente y en movimiento	Igual; hora y lugar opcionales; sincroniza con conexión
CATI	Teléfono, con ordenador en un puesto	Sí, por teléfono	Marca, registra y controla las cuotas
CAWI	Web, en el dispositivo de quien responde	No	Autocumplimentado; obligatorias y rangos

Ventajas comunes. Los saltos se aplican sin errores, no hay que grabar los datos después, se controlan los valores imposibles (un 7 en una escala de 1 a 5) y se mide la duración de cada entrevista. Inconvenientes: hay que programar y probar antes el cuestionario, y un fallo se repite en todas las entrevistas (de ahí probar todas las rutas, apartado 6.4).

En el estudio de kits. CAWI para los 78 hogares con correo, con un formulario en línea; MCAPI en la tienda, con la tableta de la zona de kits, para quien recoge su pedido y no tiene correo. Adrián encuesta en el mostrador por la tarde; el formulario es el mismo, así que los datos llegan a la misma hoja de cálculo con una columna «modo». Esa columna permite comparar respuestas por canal, sin atribuir al modo lo que puede deberse al perfil.

EN TU SECTOR

Marketing y publicidad: los estudios de audiencia de una campaña combinan CAWI rápido con CATI para llegar a quien no navega. **Gestión de ventas y espacios comerciales:** la MCAPI a la salida de una tienda o en una feria mide la experiencia de compra en el momento, con hora y lugar registrados.

ACTIVIDADES

3. Asocia cada modalidad asistida a su dispositivo, su entrevistador y su flujo de captura, y di cuál usarías para encuestar a los 34 hogares que solo tienen teléfono.
4. Explica qué ventaja tiene la MCAPI sobre una encuesta en papel en el mostrador si a la tercera pregunta hay un filtro de compra.

6.3 Redacción y escalas

Una pregunta mal escrita da un dato limpio y falso. Antes de preguntar, también hay que decir quién pregunta y para qué.

Partes del cuestionario. Ya conoces las cuatro partes: presentación, filtro, cuerpo y clasificación. En grado superior, cada parte tiene su comprobación: la presentación informa y pide el consentimiento; el filtro decide la ruta; el cuerpo mide los objetivos; la clasificación recoge solo los datos de perfil necesarios. Cuantas menos preguntas, mejor: cada una tiene que servir a un objetivo del estudio.

Reglas de redacción. Una pregunta mide **una sola idea**, con un **periodo definido** y en palabras de quien responde:

Error	Ejemplo	Versión corregida
Doble	«¿Le parecen el precio y el sabor adecuados?»	Una pregunta para el precio y otra para el sabor
Sugestiva	«¿No cree que el reparto es muy cómodo?»	«¿Cómo valora la comodidad del reparto?»
Sin periodo	«¿Compra comida preparada a menudo?»	«En las últimas cuatro semanas, ¿cuántas veces…?»
Opciones desequilibradas	«Excelente · muy bueno · bueno · regular»	Dos positivas, un punto medio y dos negativas

Además, las opciones de una pregunta cerrada son **exhaustivas** (todos encuentran la suya, con «otro» si hace falta) y **excluyentes** (nadie encaja en dos).

Escalas. Una **escala de valoración** convierte una opinión en una posición ordenada. La de tipo **Likert** pide el grado de acuerdo con una frase (de «muy en desacuerdo» a «muy de acuerdo»); la **numérica** pide un número entre dos extremos con sus etiquetas (1 nada saludable · 5 muy saludable); el **diferencial semántico** sitúa el producto entre dos adjetivos opuestos (caro · barato). Una escala equilibrada tiene tantas opciones positivas como negativas; con cinco opciones hay un punto medio, y con cuatro se obliga a elegir lado (se decide por el objetivo, no por costumbre). Ejemplo: «Los kits de La Despensa me ahorran tiempo entre semana»: muy en desacuerdo · en desacuerdo · ni de acuerdo ni en desacuerdo · de acuerdo · muy de acuerdo · no sabe. Las cinco opciones se codifican de 1 a 5 y «no sabe», con 99; la media se calcula solo con las respuestas de 1 a 5.

«No sabe» y «no procede». No sabe es una respuesta válida: quien no ha visto el precio no puede valorarlo, y obligarle a elegir crea datos falsos. **No procede** no se ofrece como opción: lo produce la **ruta**, cuando un filtro hace saltar una pregunta que no corresponde a esa persona. En la hoja de datos se codifican distinto (por ejemplo, 97 = no contesta, 98 = no procede y 99 = no sabe) y ninguno es un cero.

Información y consentimiento de quien responde. Una encuesta que recoge datos personales (un correo o un teléfono) los trata, y quien responde tiene derecho a saber quién pregunta, para qué y qué derechos tiene. La presentación del cuestionario lo dice de forma breve, en una primera capa, con un enlace o una hoja con el resto de la información. Si la base del tratamiento es el **consentimiento**, tiene que ser libre, específico, informado e inequívoco, y retirarlo tiene que ser tan fácil como darlo. El cuestionario no pide datos que no necesita (nombre o DNI) ni datos de salud que no sean imprescindibles; si un estudio lo necesitara, pediría un consentimiento explícito.

LA NORMA DICE

Ley Orgánica 3/2018, de 5 de diciembre (LOPDGDD), art. 11.2: la información básica que se da a quien facilita sus datos contiene, al menos, «a) La identidad del responsable del tratamiento […]; b) La finalidad del tratamiento; c) La posibilidad de ejercer los derechos establecidos en los artículos 15 a 22 del Reglamento (UE) 2016/679». El resto (base jurídica, plazo de conservación, reclamación ante la Agencia Española de Protección de Datos) se ofrece en una segunda capa (RGPD, art. 13). Consulta la versión vigente.

Ética de la recogida. El Código Internacional ICC/ESOMAR (edición de 2025) es un código de autorregulación, no una ley. Sus principios piden investigar con honestidad y transparencia, no perjudicar ni presionar a quien participa, explicar con claridad cómo se usan sus datos y protegerlos, y no disfrazar una venta de investigación: una encuesta no lleva una oferta del kit ni un cupón de compra (un incentivo informado sí cabe), y sus respuestas no se usan después para enviar publicidad. En la presentación de La Despensa se lee: «La Despensa de Nora, S.L. hace un estudio sobre kits de comida. Responder lleva unos 7 minutos y es voluntario. Usaremos las respuestas solo de forma agregada para decidir horarios y productos. Más información y derechos: en el enlace y en la tienda».

Hecho con aplicaciones informáticas. El cuestionario se escribe y se programa en un formulario en línea con lógica de salto (o en una hoja de cálculo con fórmulas o macros probadas): preguntas obligatorias solo donde hacen falta, rangos y una pestaña «Cambios» con versión y fecha. Las respuestas se descargan en una hoja de cálculo, con una fila por persona y una columna por pregunta.

ACTIVIDADES

5. Redacta un cuestionario neutral de diez preguntas para el estudio de kits (presentación, filtro, cuerpo y clasificación), con periodo definido, escalas equilibradas y «no sabe» donde corresponda. Prográmalo en un formulario en línea con sus saltos, ensaya cinco rutas ficticias (cliente y no cliente, hogar y oficina) y anota el tiempo de cada una y los cambios que hagas.

6. Reescribe la presentación de la encuesta si además se quisiera pedir el correo para enviar los resultados: ¿qué información añadirías y cómo separarías ese permiso del resto?

6.4 Saltos y prueba piloto

Un cuestionario no está terminado hasta que todas sus rutas se han recorrido con personas reales y un reloj.

Filtros y saltos. Un **filtro** es una pregunta que separa a quien responde en grupos (compró / no compró). Un **salto** lleva a cada grupo a la pregunta que le toca. Cada salto se escribe como regla («P7 = No → P10») y se dibuja en un **diagrama de saltos** antes de programarlo. Un salto mal programado no da ningún aviso: da respuestas de quien no debía contestar.

Diagrama de saltos del cuestionario v0.2 (estudio de kits)

Pregunta	Condición	Destino
P1-P6	Todos (hábitos de compra y de horario)	P7
P7 Filtro: compra de kits en las últimas cuatro semanas	Sí	P8
P7	No	P10
P8 Precio · P9 Saludable	Solo compradores	P10
P10-P12	Todos (intención, franja preferida, perfil)	Fin

Cada ruta es un camino completo de la primera pregunta al final. Este cuestionario tiene dos rutas por el filtro (compradores y no compradores), y con el perfil de P12 se cruza con dos tipos de cliente (hogar y oficina): cuatro perfiles que hay que probar. Una ruta falla si muestra una pregunta indebida (satisfacción a quien no compró), deja a alguien **sin salida** (un salto a una pregunta que no existe) o se salta una pregunta que sí correspondía. Para no olvidar ninguna, se prepara una **tabla de rutas**: una fila por perfil, con las preguntas que debe ver, las que no debe ver y una casilla para marcar que se ha probado.



La prueba piloto o pretest. El **pretest** aplica el cuestionario a pocas personas parecidas a las de la población, con el mismo modo de recogida, antes del trabajo de campo. No estima opiniones de la población; sirve para medir y corregir el instrumento: **duración**, **comprensión** (dudas, palabras difíciles), **coherencia** de las rutas y **opciones que faltan**. Cada persona de prueba se registra con un código y un perfil, sin nombre. ¿Cuántas? Al menos una por ruta, y mejor dos en las rutas nuevas o corregidas. Si una corrección cambia una ruta, se vuelve a probar esa ruta; si cambia la redacción de una pregunta, se prueba con alguien que no la haya visto antes.

Plantilla de pretest (una fila por persona de prueba)

Campo	Qué se anota
Código y perfil	T1, hogar cliente…
Ruta prevista y ruta real	Preguntas que debía ver y las que vio
Duración	Minutos y segundos, de la presentación al fin
Incidencias	Dudas literales, saltos erróneos, opciones que faltan
Cambio propuesto y versión	Qué se corrige y en qué versión (v0.1 → v0.2)



Registro de cambios.

Cada corrección queda escrita con la versión, la fecha, el motivo y quién la aprueba: «v0.1 → v0.2 · 23/03/2027 · filtro de compra añadido como P7, resto renumerado, tras el pretest (T2 y T4) · aprueba Nora». Así, si en mayo alguien pregunta por qué cambió una pregunta, la respuesta está en el archivo y no en la memoria de nadie.

ACTIVIDADES

7. Dibuja el diagrama de saltos de un cuestionario para oficinas con dos filtros: «¿Su empresa pide comida para reuniones?» (No → fin, tras agradecer) y «¿Quién decide la compra?» (si no es la persona que responde → preguntas sobre quién decide). Escribe todas las rutas y di cuántas personas de prueba necesitarías como mínimo para recorrerlas.

8. Aplica el pretest al conflicto de la unidad: con la plantilla, registra una ruta de un hogar no cliente con el cuestionario v0.1 y otra con el v0.2, y di qué condición te haría repetir el pretest antes de lanzar la encuesta.

6.5 Paneles y elección de técnica

Una encuesta transversal es una foto; un panel es una serie de fotos a las mismas personas. Elegir entre ellas depende de lo que hay que decidir, del plazo y del dinero.

Qué es un panel. Un **panel** es una muestra fija de unidades definidas (hogares, tiendas, personas) que informan **de forma repetida**, cada cierto tiempo y con el mismo método. Por eso permite ver cambios en las mismas unidades, algo que una **encuesta transversal** (una sola medición a una muestra) no da. No es «cualquier grupo consultado una vez».

Tipo de panel	Unidad	Qué registra y cada cuándo
De consumidores	Hogares o personas	Compras o consumo, con diario o lector, cada semana o cada mes
De detallistas	Tiendas	Ventas, existencias y precios por producto, cada mes o cada bimestre
De audiencias	Hogares o personas	Exposición a medios (televisión, radio, internet), de forma continua

Problemas propios del panel. El **desgaste** o **mortalidad** del panel es la pérdida de participantes entre oleadas, que puede no ser aleatoria (abandonan más quienes tienen menos tiempo); se compensa con reposición y se informa. El **condicionamiento** aparece cuando, por estar en el panel, la persona cambia su conducta o sus respuestas (empieza a fijarse en lo que compra). Un panel necesita incentivo y mantenimiento continuos: es caro para un estudio puntual.

Ejemplo. El desgaste de un panel. Un panel de 40 tiendas de alimentación informa cada mes de sus ventas de comida preparada. En enero informan 40; en febrero, 37, y en marzo, 34. El desgaste acumulado es 6 de 40 = **15,0 %**. Si las que abandonan son las más pequeñas, las ventas medias del panel suben aunque el mercado no cambie: por eso se reponen con tiendas parecidas (anotadas aparte), los cambios se miden en las que continúan y el desgaste se publica.

Otras opciones. Una **encuesta ómnibus** es una encuesta periódica de una empresa de investigación en la que varios clientes compran unas pocas preguntas: es más barata que un estudio propio, pero el cuestionario es corto y la población no la eliges tú. Para la decisión de Nora (qué franja de recogida ofrecer) basta una medición, es decir, una encuesta transversal; un panel tendría sentido si quisiera seguir mes a mes las compras de los mismos hogares después del cambio de horario.

Elegir con recursos reales. Nora tiene **cuatro semanas** de abril (05/04 a 30/04/2027) y como máximo 100,00 € para la recogida. Las opciones:

Cuadro 1 · Opciones de recogida

(euros sin IVA; horas del equipo)

Opción	Cobertura máx. (%)	Desembolso (€)	Horas
A · Solo en línea (CAWI)	65,0	0,00	6
B · En línea + MCAPI en tienda	95,0	0,00	22
C · En línea + CATI	93,3	0,00	15
D · Postal con respuesta pagada	53,3	96,00	8

B suma a las 6 horas del formulario 8 tardes de 2 horas (16 h); C, unos 15 minutos por llamada a los 34 hogares con solo teléfono (8,5 h, que con la preparación se quedan en 9 h). D cuesta 64 envíos × 1,50 € = 96,00 € más 8 h (6 de impreso, 2 de envío y grabación); su plazo de vuelta apenas cabe en cuatro semanas. La de menor desembolso, A, deja fuera al 35,0 % y casi a todos los mayores; B es la que mejor responde al objetivo (conocer la franja preferida de los compradores) con el presupuesto y el plazo, aunque pide más horas y su cobertura depende de que pasen por el mostrador y contesten. Lo que sigue fuera (6 hogares sin correo que no recogen en tienda) se dice en el informe.

Calendario.

Enlace el lunes 05/04 y recordatorio el lunes 12/04; mostrador del martes 06/04 al jueves 29/04; cierre el viernes 30/04.

CLAVE · BARATO NO ES ADECUADO

La técnica se elige por el objetivo, la población a la que hay que llegar, el plazo y el presupuesto, en ese orden. Si la opción no llega a todos, el informe lo dice.

ACTIVIDADES

9. Elige modalidad para el encargo de cuatro semanas de Nora y justifica una alternativa descartada con datos del Cuadro 1. ¿Qué cambiaría tu elección si el presupuesto fuera de 0,00 € y Adrián solo pudiera estar dos tardes?

10. Compara los tres tipos de panel con una encuesta transversal (unidad, periodicidad, coste, desgaste y condicionamiento) y diseña un seguimiento breve para el estudio de kits: 20 hogares compradores que anotan sus compras de kits durante cuatro semanas.

CASO RESUELTO · PASO A PASO

La pregunta del kit «excelente»

Caso ficticio. Jueves 18 de marzo de 2027. El equipo corrige la pregunta 8 del cuestionario v0.1 y lo prueba con cinco personas (T1-T5) en la zona de kits. Adrián cronometra y anota las rutas.

Paso 1. Decisión, datos y alcance. Decisión: lanzar o no el cuestionario el 05/04. Pregunta de partida: «¿No cree que nuestro excelente kit es barato y saludable?». Alcance: comprobar el instrumento, no estimar opiniones.

Paso 2. Diagnóstico y corrección. La pregunta es **sugestiva** («¿No cree…?», «excelente») y **doble** (precio y salud a la vez); no tiene periodo ni filtro de compra, y solo admite sí o no. Versión v0.2: P7 «En las últimas cuatro semanas, ¿ha comprado algún kit de La Despensa? Sí / No» (No → P10); P8 «El precio del último kit que compró le pareció: muy caro · algo caro · adecuado · algo barato · muy barato · no sabe»; P9 «¿Cómo valora lo saludable de ese kit? (1 nada saludable · 5 muy saludable) · no sabe».

Paso 3. Pretest con el cuestionario v0.1

Persona	Perfil	Duración	Incidencia
T1	hogar cliente	6 min 40 s	ninguna
T2	hogar no cliente	8 min 10 s	le aparece la satisfacción: falta el filtro
T3	oficina cliente	5 min 20 s	ninguna
T4	oficina no cliente	9 min 30 s	falta el filtro; duda en P8
T5	hogar cliente	7 min 00 s	P8 aún mezcla precio y salud

Duración media: (400 + 490 + 320 + 570 + 420) ÷ 5 = 2.200 ÷ 5 = 440 s = **7 min 20 s**; máximo 9 min 30 s. Fallan las dos rutas de no cliente (T2, T4): con la v0.2 se repiten el martes 23/03: T2 tarda 5 min 10 s y T4, 5 min 40 s, y ninguna ve P8 ni P9; las de comprador siguen sin probar.

Paso 4. Decisión y registro. Nora aprueba el v0.2 (registro del 23/03) y fija el lanzamiento para el lunes 05/04, si antes se prueban las rutas de comprador con personas nuevas, también en la tableta del mostrador (el modo de los hogares sin correo). La presentación dirá «unos 7 minutos» como estimación (v0.1: 7 min 20 s; el v0.2 es más corto para quien no compra) hasta medirlo.

RESULTADO

La pregunta sugestiva y doble se sustituye por dos neutrales (precio y salud), con opciones equilibradas, «no sabe», periodo definido y filtro (el «no procede» lo da la ruta). El pretest permite concluir que el v0.2 corrige las dos rutas de no cliente; faltan las de comprador y medir su duración (7 minutos, estimada). No permite concluir nada sobre lo que opinan los clientes: cinco personas de prueba no son una muestra.

PRÁCTICA

El cuestionario de Ruedalia

Txema quiere encuestar a trabajadores del polígono para su servicio a empresas. El cuestionario v0.1 tiene seis preguntas: Q1 «¿Viene al trabajo en bicicleta? Sí / No»; Q2 «¿Cuántas veces ha llevado su bicicleta a Ruedalia en el último año?»; Q3 «¿Qué tal fue el servicio de Ruedalia? (1-5)»; Q4 «¿Recomendaría Ruedalia a un compañero?»; Q5 «¿Le interesaría una revisión en su empresa?»; Q6 edad por tramos. Ousmane lo ha probado con cuatro personas:

Persona	Perfil	Duración	Incidencia
C1	cliente	3 min 10 s	ninguna
C2	va en bici, nunca ha ido a Ruedalia	4 min 20 s	contesta Q3 y Q4 «por contestar»
C3	no va en bici	4 min 50 s	contesta todo; duda en Q2
C4	cliente	2 min 40 s	ninguna

Encargo. 1) Añade los filtros y saltos necesarios y dibuja el diagrama. 2) Calcula la duración media del pretest. 3) Indica qué rutas hay que volver a probar y con qué perfiles. 4) Decide si el cuestionario se puede lanzar y con qué limitaciones.

Plantilla IC-P06 · Cuestionario de Ruedalia

Campo	Tu respuesta
Datos y fuentes con unidad y periodo	
Procedimiento o cálculo	
Resultado y evidencia	
Decisión y limitaciones	

Lista de cotejo

☐ Ninguna ruta pregunta satisfacción a quien no es cliente.

☐ Cada salto tiene destino.

☐ La duración se da con su unidad.

☐ Se dice qué no prueba el pretest.

LECTURA Y DEBATE

Coste cero



Relato ficticio. El martes 16 de marzo, sin esperar al pretest, Unai mandó una versión de prueba del formulario a los 78 hogares con correo. «Para ir viendo», dijo. El jueves por la mañana tenía 41 respuestas y una sonrisa: «El 52,6 % de respuesta en dos días. Ninguna encuesta en papel hace eso».

Adrián abrió la hoja y, como siempre, contó antes de opinar. De las 41 respuestas, solo 2 eran de personas de 65 años o más: un 4,9 %. Luego miró el registro de Itxaso: entre los 120 hogares compradores, 36 tienen 65 años o más, el 30,0 %. «Los que más kits recogen por la tarde casi no salen en la encuesta», le dijo a Olvido, que asintió sin sorpresa.

Unai defendió su decisión: muchas respuestas, rápidas y sin coste. Nora le dio la razón en lo rápido, pero no en lo demás: «Me has traído muy bien la opinión de una parte. Yo necesito la de todos los que compran». La comprobación no fue una opinión de Olvido, sino comparar el perfil de quien respondió con el del registro.

Esa tarde decidieron que el formulario en línea se quedaba, pero acompañado: la tableta en el mostrador para quien no tiene correo. Y las 41 respuestas de prueba no se mezclarían con las de abril, porque se hicieron con un cuestionario sin probar.

Preguntas. 1) ¿Qué afirmación del caso necesita una comprobación antes de convertirse en decisión? 2) ¿Qué información pedirías primero y por qué cambiaría la respuesta? 3) Propón una actuación provisional que permita aprender sin ocultar la incertidumbre.

APRENDE CON LA IA

Unai pregunta a un asistente de IA si le conviene «montar un panel» para el estudio de kits. La respuesta empieza así:

RESPUESTA DE LA IA

«Un panel es una encuesta realizada una sola vez a cualquier grupo.»

Tu tarea. Localiza el error, contrástalo con lo que has aprendido en la unidad y escribe una respuesta corregida con su límite (qué exige un panel y qué problemas tiene). Conserva la definición o la fuente que justifica la corrección.

Ojo con la IA. Una definición dicha con seguridad no es una definición correcta: compárala con la de un manual o con la de un organismo de estadística.

ACTIVIDADES FINALES

Comprueba

11. Resume la distinción entre cobertura y tasa de respuesta y aplícala al caso resuelto y a la lectura.
12. Explica con un ejemplo del caso qué hace un filtro y qué condición te obligaría a repetir el pretest.
13. ¿Qué diferencia hay entre «no sabe» y «no procede»? ¿Cómo se codifica cada uno?

Aplica

14. Repite la práctica de Ruedalia cambiando una condición: el cuestionario se hará por teléfono (CATI) a 40 clientes de empresas. Señala qué resultado cambia (duración aceptable, forma de las escalas, quién aplica los saltos) y cuál se mantiene (los filtros).
15. Calcula la duración media del pretest del caso si T2 y T4 hubieran tardado lo que tardaron con el v0.2 (5 min 10 s y 5 min 40 s). ¿Por qué no debes dar esa media como duración definitiva?
16. Redacta la presentación de la encuesta de Ruedalia con la información básica para quien responde.

Investiga

17. Localiza una fuente primaria o una documentación técnica sobre paneles y elección de técnica (la metodología de una encuesta del Instituto Nacional de Estadística, la explicación de un panel de audiencias o de consumidores de una asociación del sector). Registra fecha, ámbito, una afirmación que sustenta lo que has aprendido y una limitación.
18. Busca en la web de la Agencia Española de Protección de Datos un modelo o una guía de información por capas y compara su primera capa con la presentación de La Despensa.

TU INVESTIGACIÓN AVANZA · PASO 6 · I06 · CUESTIONARIO PROBADO

Para tu negocio, redacta el cuestionario con presentación informativa, filtro, cuerpo y clasificación; prográmalo en un formulario en línea u hoja de cálculo; dibuja el diagrama de saltos y haz un pretest con al menos cuatro personas que recorran todas las rutas. Anota duración, incidencias y cambios con versión y fecha, y justifica la modalidad elegida con su cobertura. Incluye la decisión, el responsable y la fecha de revisión. Entrega el archivo del formulario y el registro del pretest, no solo una presentación.
Comprueba tu entrega: ninguna pregunta es doble ni sugestiva; ninguna ruta pregunta algo indebido; la presentación dice quién pregunta, para qué y qué derechos hay.
☐ Hecho ☐ Revisado ☐ Guardado

OJO PROFESIONAL

Documento: cuestionario EK-2027/CU-v0.1 (extracto)

Documento ficticio de La Despensa de Nora, S.L., preparado por Unai el lunes 15 de marzo de 2027. Contiene **cinco errores**.

N.º	Pregunta o nota	Texto del borrador
1	Presentación	«La Despensa de Nora, S.L. hace un estudio sobre kits para decidir horarios y productos. Es voluntario y dura unos 7 minutos. Información y derechos: en el enlace»
2	P3	«¿Compra usted comida preparada a menudo? Sí / No»
3	P7	«¿Qué tal su experiencia con nuestros kits? (1 muy mala · 5 muy buena)», para todas las personas
4	P8	«¿No cree que nuestro excelente kit es barato y saludable? Sí / No»
5	P11	«Franja preferida para recoger: 13:00-13:59 · 17:00-17:59 · 18:00-18:59 · 19:00-19:59 · otra · no sabe»
6	P12	«Valore la atención: excelente · muy buena · buena · regular»
7	Nota de programación	«Saltos programados en el formulario. No hace falta probarlos: el programa no se equivoca»

Encargo. Señala los cinco errores, explica por qué cada uno es un error con lo que has aprendido en la unidad y reescribe las preguntas afectadas para que el cuestionario se pueda utilizar. Di qué dos líneas están bien.

N.º	Línea	Por qué es un error	Corrección
1			
2			
3			
4			
5			

AUTOEVALUACIÓN

1. De 120 hogares, 78 tienen correo; de los 36 de 65 años o más, solo 6. Una encuesta solo en línea...

a) cubre el 65,0 % y deja fuera a casi todos los mayores.
b) cubre a todos si responden bastantes.
c) no tiene sesgo.

2. Adrián encuesta en el mostrador con la tableta, que aplica los saltos y envía la respuesta con hora. Es...

a) CATI.
b) MCAPI.
c) CAWI.

3. «¿Le parecen el precio y el sabor adecuados?» es...

a) correcta si tiene «no sabe».
b) doble: hay que separarla en dos preguntas.
c) sugestiva, pero no doble.

4. En el pretest, a un no cliente le aparece la pregunta de satisfacción. ¿Qué haces?

a) Lanzar y borrar luego esas respuestas.
b) Corregir el salto, anotar el cambio y repetir esa ruta.
c) Añadir «no procede» como opción.

5. Cuatro semanas, 100,00 € y un marco con muchos mayores sin correo. ¿Qué eliges?

a) Solo en línea, porque cuesta 0,00 €.
b) En línea más MCAPI en tienda, informando de lo que queda fuera.
c) Postal, porque llega a todos.

6. La media del pretest del caso (440 s) es...

a) 7 min 20 s, y no dice nada de la opinión de los clientes.
b) 7,4 min, y estima la satisfacción.
c) 7 min 33 s.

7. En Ruedalia, un trabajador que nunca ha ido al taller contesta «qué tal fue el servicio». Se corrige con...

a) una escala de 1 a 10.
b) un filtro en Q2 que salte Q3 y Q4 a quien no ha ido nunca.
c) borrar Q3.

8. «Un panel es una encuesta realizada una sola vez a cualquier grupo.» ¿Qué es correcto?

a) Es correcto si el grupo es grande.
b) Un panel mide de forma repetida a unidades definidas.
c) Un panel es un grupo de discusión.

9. En el cuestionario v0.1, «No hace falta probar los saltos: el programa no se equivoca».

a) Correcto.
b) Error: un fallo de programación se repite en todas las entrevistas; se prueban todas las rutas.
c) Basta la ruta de clientes.

10. Tu I06 usa solo un formulario en línea. ¿Qué te haría revisar la modalidad?

a) Ninguno.
b) Que el perfil de quien responde sea muy distinto del perfil de tu marco (por edad o por canal).
c) Los colores.

Ya sé...

- ☐ explicar las distinciones de los cinco apartados: modalidades, siglas, redacción, saltos y paneles.
- ☐ ejecutar el procedimiento del caso: corregir una pregunta doble y sugestiva y hacer un pretest con duración y rutas.
- ☐ transferirlo a otro negocio, como el cuestionario de Ruedalia.
- ☐ detectar una conclusión no sustentada, como «coste cero» sin mirar la cobertura.
- ☐ entregar una evidencia revisable: formulario, diagrama de saltos y registro de pretest.



UNIDAD 7

Muestreo y diseño de selección

Bloque 3 · Muestreo y preparación de datos

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

Lunes 5 de abril de 2027. El cuestionario del Estudio de kits 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;
- cómo se elige cada hogar para poder generalizar, y hasta dónde.

PARA EMPEZAR

1. ¿Qué debe decidir Nora: cuántas respuestas quiere o a quién quiere poder describir con ellas?
2. ¿Qué sabemos de quien sale de la tienda una mañana, y qué suponemos de quien no sale?
3. Si preguntas a cien personas en la puerta, ¿a qué población representan esas cien respuestas?
4. ¿Qué prueba permitiría revisar la primera respuesta de Nora antes de salir a la calle?



Población y marco

7.1



Aleatorio y sistemático

7.2



Estratos y conglomerados

7.3



No probabilísticos

7.4



Sesgos y protocolo

7.5

7.1 Población y marco

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

Población objetivo. Ya distingues población, muestra y muestreo; ahora hay que escribirlo todo. La **población objetivo** es el conjunto de unidades sobre el que la decisión necesita información, definido por su contenido (qué son), su ámbito (dónde) y su tiempo (cuándo). Para el Estudio de kits: *hogares que viven en el barrio y oficinas del barrio y del polígono, en abril de 2027*. **Universo** se usa a menudo como sinónimo.

Unidad de muestreo, de análisis e informante. La **unidad de muestreo** es lo que se sortea; la **unidad de análisis**, aquello de lo que se obtienen los datos; el **informante**, quien contesta. En hogares se sortea el hogar y responde quien hace la compra; en oficinas se sortea la empresa y contesta quien propone las compras, como Mencía en Albada Asesores. Sortear tiques en lugar de hogares daría más peso a quien compra más veces.

Marco muestral. El **marco** es la lista, el mapa o el registro del que se seleccionan las unidades. Solo lo que está en el marco puede salir, así que el marco define la **población de estudio**: la que de verdad se puede describir. Antes de sortear se revisan:

- **Subcobertura**: unidades de la población que no están en el marco (los hogares que nunca han dado sus datos a La Despensa).
- **Sobrecobertura o elementos ajenos**: unidades del marco que no son de la población (una oficina en una lista de hogares).
- **Duplicados**: una unidad que aparece dos veces y tiene más probabilidad de salir.

Tabla 1 · Población objetivo y marcos posibles del Estudio de kits

(ficticia)

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.

Fases del proceso de muestreo. Todo muestreo sigue el mismo orden, y cada fase deja un registro:

1. Definir la población objetivo y la unidad de muestreo.
2. Elegir o construir el marco y depurarlo.
3. Elegir el método de selección (7.2-7.4).
4. Determinar el tamaño de la muestra (unidad 8).
5. Seleccionar las unidades y hacer el trabajo de campo con registro de contactos.
6. Revisar cobertura, sustituciones y no respuesta, y documentarlo (7.5).

Depurar el marco antes de sortear. El martes 6 de abril, Adrián revisa la **lista de interés de kits**: hogares que pidieron información de los kits en la tienda o en la web entre el 11 de enero y el 31 de marzo y autorizaron que se les contactara para estudios. Busca teléfonos y correos repetidos y anota lo que encuentra; el equipo decide y lo deja escrito:

Registro 1 · Depuración de la lista de interés de kits

(ficticio)

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

Control: 104 − 3 − 1 = 100. El archivo original se conserva sin tocar y se trabaja sobre una copia con versión.



LA NORMA DICE

Reglamento (UE) 2016/679 (Reglamento general de protección de datos), de 27 de abril de 2016, art. 5.1 c): los datos personales serán «adecuados, pertinentes y limitados a lo necesario en relación con los fines para los que son tratados («minimización de datos»)». Para sortear basta código, zona y contacto autorizado. Consulta la versión vigente.

ACTIVIDADES

1. Escribe, para el Estudio de kits, la población objetivo de las oficinas: contenido, ámbito y tiempo. Indica la unidad de muestreo, la unidad de análisis y el informante, y explica qué cambiaría si se sorteara a las personas que trabajan en cada oficina.
2. Aplica 7.1 al conflicto de la unidad: elige un marco para hogares de la tabla 1 (o propón otro), documenta con un ejemplo qué subcobertura y qué duplicados tendría, da una alternativa y escribe la condición que te haría cambiar de marco.

7.2 Aleatorio simple y sistemático

Una muestra es probabilística cuando cada unidad del marco tiene una probabilidad conocida y mayor que cero de salir, y la elige un procedimiento, no una persona.

Probabilidad de inclusión. Es la probabilidad que tiene cada unidad del marco de entrar en la muestra. Si se conoce para todas, se puede estimar con un margen de error y saber a qué población se refiere el resultado. Si no se conoce, la muestra puede servir para explorar, pero no para calcular un margen de error.

Muestreo aleatorio simple sin reposición (m.a.s.). Se eligen n unidades de un marco de N de forma que **todas las muestras posibles de tamaño n tengan la misma probabilidad**; sin reposición significa que una unidad elegida no puede volver a salir. Cada unidad tiene una probabilidad de inclusión n/N: con N = 100 y n = 10, 10/100 = 0,10; y hay 17.310.309.456.440 muestras distintas posibles. Ventaja: es la base de las fórmulas de la unidad 8. Inconvenientes: exige un marco completo y numerado, y la muestra puede quedar muy dispersa.

Cómo sortear sin repetir y dejar rastro. En una hoja de cálculo (Google Sheets, LibreOffice Calc u otra):

1. Columna A: identificadores del marco (H001 a H100). Columna B: =ALEATORIO() (en inglés, =RAND()) en cada fila.
2. Copiar la columna B y pegar **solo los valores**, para que no cambien al recalcular.
3. Ordenar por la columna B, tomar las n primeras y guardar el archivo con fecha, versión y autor.

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

Muestreo sistemático con arranque aleatorio. Se calcula el **intervalo** k = N/n, se sortea un **arranque** a entre 1 y k y se toman a, a + k, a + 2k… hasta completar n. Es rápido y reparte la muestra a lo largo de la lista. Cada unidad tiene la misma probabilidad de inclusión que en el m.a.s. (1/k), pero solo hay **k muestras posibles**: dos unidades vecinas nunca pueden salir juntas. Por eso **no es un aleatorio simple**, aunque se le parezca.

La periodicidad del marco. El sistemático falla si la lista tiene un ciclo que coincide con k o con un divisor de k. Contraejemplo: si el marco estuviera ordenado por portales de diez viviendas, con k = 10 y arranque 7 saldría siempre la séptima vivienda de cada portal, que podría ser siempre la misma planta. Antes de sortear se mira cómo está ordenada la lista; si hay un ciclo, se reordena (por ejemplo, por fecha de alta) o se usa el m.a.s.

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.

ACTIVIDADES

3. Clasifica estos métodos según la probabilidad de inclusión y la posibilidad de inferir a la población: a) los 30 primeros clientes de la mañana; b) 10 hogares del marco v1.0 sorteados con semilla; c) uno de cada diez hogares del marco empezando por el primero; d) uno de cada diez con arranque sorteado; e) 20 entrevistas con 10 hombres y 10 mujeres elegidos por el entrevistador.

4. Selecciona en una hoja de cálculo 8 identificadores sin reposición de un marco H001 a H040 con el procedimiento de los tres pasos. Registra la fecha, la versión del marco y el archivo, y explica cómo comprobarías que no hay ninguno repetido.

7.3 Estratificado y conglomerados

Estratificar es dividir para que no falte ningún grupo; conglomerar es agrupar para que el trabajo de campo sea posible. Se parecen en el dibujo y son opuestos en la lógica.

Muestreo estratificado. Se divide el marco en **estratos**: grupos que no se solapan, que juntos cubren todo el marco y que son internamente parecidos respecto a lo que se estudia. Después se sortea dentro de **cada** estrato. Garantiza que ningún grupo falte y gana precisión si los estratos son distintos entre sí. Requisito: conocer el estrato de **cada** unidad del marco antes de sortear; si la zona de un hogar no está en la lista, ese hogar no se puede estratificar. Los estratos se definen con variables que se sabe que influyen en lo que se mide: zona, tipo de hogar u oficina, antigüedad como cliente. En el marco depurado, cada hogar lleva su zona: Norte 40, Centro 36 y Sur 24. Una muestra de 10 proporcional a esos tamaños tendría 4 hogares del Norte, 4 del Centro y 2 del Sur (las cuotas exactas, 4; 3,6 y 2,4, se redondean por mayores restos). El cálculo del reparto entre estratos se trabaja a fondo en la unidad 8.

Muestreo por conglomerados. Un **conglomerado** es un grupo natural de unidades (una empresa con su plantilla, una manzana con sus viviendas) que se sortea **entero**. Interesa cuando hay lista de grupos pero no de unidades, o cuando recorrer unidades dispersas sería carísimo. Conviene que cada conglomerado sea variado por dentro; si sus unidades se parecen mucho, cada entrevista aporta menos y hace falta más muestra. Ejemplo: de un directorio de 30 empresas del polígono se sortean 5 y se entrevista a todas las personas que comen en la empresa ese día.

Por áreas y por etapas. Cuando los conglomerados son zonas del mapa (secciones, manzanas, calles) se habla de **muestreo por áreas**. Si dentro de cada grupo elegido se vuelve a sortear, el muestreo es **polietápico**: primero manzanas, después viviendas dentro de cada manzana y, por último, una persona dentro de cada vivienda. Cada etapa tiene su marco y su sorteo; la probabilidad final es el producto de las de cada etapa.

Ruta aleatoria. El entrevistador parte de un punto sorteado y sigue unas **reglas fijas** de recorrido (girar a la derecha, uno de cada tres portales, la persona que cumpla años antes). Solo es probabilística si la selección de puntos, hogares y persona y las sustituciones están documentadas y dan probabilidad de inclusión conocida. Si el entrevistador puede saltarse portales o cambiar de calle, es conveniencia.

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

Un ejemplo real de etapas. El INE describe su encuesta de tecnologías en los hogares (TIC-H) de 2025 como un muestreo **trietápico estratificado**: secciones censales, después viviendas familiares principales y, por último, una persona de 16 o más años de cada vivienda. Su tamaño muestral teórico fue de 26.862 viviendas (INE, nota de prensa de la TIC-H 2025, consultada el 28/10/2026).

ACTIVIDADES

5. Resuelve el sistemático del marco de 100 hogares con intervalo 10 y arranque 7, escribe los identificadores y explica por qué el orden por fecha de alta es preferible al orden por portal. Después, diseña en una tabla los esquemas de estratos, conglomerados, áreas, etapas y ruta para las oficinas del polígono: qué se sortea en cada uno y con qué lista.

6. Itxaso dice que «estratificar por zona y sortear manzanas es lo mismo». Explica la diferencia con un ejemplo de cada uno y di cuál elegirías para que no falte ninguna zona en la muestra.

7. Redacta las reglas de una ruta aleatoria para entrevistar a 12 hogares del barrio: punto de partida, recorrido, vivienda, persona y sustituciones. Señala qué regla impide que la ruta se convierta en conveniencia.

7.4 Conveniencia, cuotas y bola de nieve

Las selecciones no probabilísticas son rápidas y a veces las únicas posibles. Sirven para explorar y probar; no sirven para calcular un margen de error.

Conveniencia. Se pregunta a quien está a mano: los que salen de la tienda, los seguidores de las redes, los compañeros de clase. Es barata e inmediata, pero nadie conoce la probabilidad de cada persona. La propuesta de Nora es una muestra de conveniencia: una mañana en la puerta deja fuera a quien trabaja a esa hora, a las oficinas, a quien compra solo por la web y a todos los que nunca entran.

Juicio (u opinática). Una persona experta elige a quienes considera típicos o informativos. Olvido podría escoger a cinco habituales «que saben de cocina» para probar el cuestionario o los kits. Es útil para un pretest o para elegir informantes clave en una fase exploratoria; pero refleja lo que Olvido cree típico, y sus habituales no son los hogares que no compran. Su resultado se presenta como opinión de informantes elegidos, no como porcentaje de la población.

Cuotas. Se fija cuántas personas de cada grupo (edad, tipo de cliente, zona) hay que conseguir, a menudo en la proporción que tienen en la población, y el entrevistador **elige libremente** a quién entrevistar dentro de cada cuota. La muestra se parece a la población en las variables de cuota, pero dentro de cada casilla se cuela la conveniencia: el entrevistador encuentra antes a quien está en la calle, a quien no tiene prisa, a quien se le parece. Que la muestra cumpla las cuotas no hace que cada persona tenga una probabilidad conocida.

Ejemplo. Unai propone 20 entrevistas en la puerta con cuotas: 10 hogares y 10 oficinas, y en cada grupo 5 personas de menos de 45 años y 5 de 45 o más. Por la mañana no se llenan las cuotas de oficina; a mediodía las llena quien recoge su pedido. Quien nunca viene sigue sin estar.

Bola de nieve. Cada participante propone a otros que cumplen el perfil. Sirve para poblaciones sin lista o difíciles de encontrar: Mencía podría presentar al equipo a otras responsables de servicios generales del barrio. Riesgo: un círculo de conocidos parecidos.



Para qué sí sirven. Explorar motivos y lenguaje, probar un instrumento o encontrar casos raros. No para dar porcentajes de toda la población ni márgenes de error.

Cómo se informa de una selección no probabilística. El informe dice cómo se eligió a las personas, dónde y cuándo; da los resultados como recuentos y porcentajes **de quienes respondieron** («24 de las 30 personas entrevistadas a la salida»), sin margen de error; y nombra a quién deja fuera. Si después hay una muestra probabilística, se comparan las dos y se explica la diferencia.

CLAVE · CUOTAS CUMPLIDAS NO ES LO MISMO QUE MUESTRA PROBABILÍSTICA

Las cuotas controlan la composición en unas pocas variables; no controlan quién entra. Sin probabilidad de inclusión conocida no hay margen de error, por bien que cuadren las casillas.

ACTIVIDADES

8. Diseña cuatro selecciones ficticias para el Estudio de kits (una por conveniencia, otra por juicio, otra por cuotas y otra por bola de nieve). Para cada una, explica el mecanismo de selección y qué población queda infrarrepresentada.
9. Unai dice que sus 20 entrevistas con cuotas «representan el barrio igual que un sorteo». Contesta en cinco líneas qué controla su diseño, qué no controla y qué se podría afirmar con sus resultados.
10. ¿En qué momento del Estudio de kits elegirías una selección por juicio, y por qué no la usarías para estimar qué proporción de hogares compraría un kit?

7.5 Sesgos y selección documentada

Un tamaño grande no corrige una selección torcida. Lo que no está escrito no se puede revisar.

Error de muestreo y sesgo. El **error de muestreo** es la diferencia que aparece porque se observa una muestra y no toda la población; en un diseño probabilístico se puede calcular y se reduce al aumentar la muestra. Un **sesgo** es una desviación sistemática que empuja el resultado siempre hacia el mismo lado y que **no** baja al aumentar la muestra. Un informe honesto da el primero con su cálculo y nombra el segundo con su causa probable, aunque no pueda medirlo. Ejemplo: «error de muestreo de ±X puntos; posible sesgo de cobertura: la lista no incluye a quien nunca pidió información».

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

Sesgo	Qué ocurre	Ejemplo
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 de La Despensa reunió 286 respuestas en una semana; 241 eran de personas que ya compraban en la web (84,3 %). Describen a la clientela digital, no al barrio. Diez hogares bien sorteados dicen menos, pero se sabe a quién se refieren.

El protocolo de selección. El diseño se escribe **antes** del campo y lo que pasa **durante** el campo se registra.

Documento 1 · Protocolo de selección EK-2027/PS-v1.0

(ficticio)

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

Sustituciones y no respuesta. Si un hogar seleccionado no contesta, la tentación es llamar al siguiente de la lista. Sin regla, la muestra se desplaza hacia los hogares fáciles. Lo correcto: hacer los intentos previstos en horarios distintos, registrar cada resultado (completa, rechazo, no localizado, fuera del marco) y, si se sustituye, hacerlo con una **reserva sorteada** a la vez que la muestra y en el orden sorteado. La tasa de respuesta y su efecto se estudian en la unidad 8.

Comparar métodos por coste y límites. Para hogares con lista, un sistemático telefónico es barato y permite inferir a la lista. Para oficinas pequeñas sin directorio fiable, un muestreo por áreas (recorrer manzanas y registrar las oficinas) cuesta más horas; una bola de nieve desde Albada Asesores es barata, pero no permite estimar porcentajes.

ACTIVIDADES

11. Selecciona un método para los hogares y otro para las empresas de difícil acceso del polígono. Compara en una tabla el coste (horas, desplazamientos, llamadas), el límite principal y lo que permitiría concluir cada uno.
12. Explica con la encuesta de la web por qué 286 respuestas no son representativas aunque sean muchas. ¿Qué dato pedirías para saber a quién representan?
13. Añade al protocolo EK-2027/PS-v1.0 la regla de sustitución con reserva sorteada: cuántos hogares de reserva, cómo se sortean, cuándo se usan y cómo se registra.

CASO RESUELTO · PASO A PASO

El sistemático de la lista de interés

Caso ficticio. El jueves 8 de abril de 2027 el equipo selecciona 10 de los 100 hogares del marco depurado para una entrevista telefónica sobre la entrega programada de kits.

Paso 1. Decisión, datos y alcance. Decisión: qué hogares se llaman, sin que nadie los elija a mano. Datos: marco v1.0, 100 hogares numerados H001-H100 por fecha de alta, sin duplicados; n = 10. Alcance: el resultado describirá a los hogares de la lista de interés, no al barrio ni a los que nunca han pedido información.

Paso 2. Procedimiento. Intervalo: k = N/n = 100/10 = **10** hogares. Arranque: un número entero sorteado entre 1 y 10 con =ALEATORIO.ENTRE(1;10) en la hoja, pegado como valor y anotado en el protocolo: salió **7**. Selección: 7, 7 + 10 = 17, 27, 37, 47, 57, 67, 77, 87 y 97 (el siguiente, 107, ya está fuera del marco).

Posición	Identificadores	Zona
7, 17, 27, 37, 47	H007, H017, H027, H037, H047	C, S, C, N, C
57, 67, 77, 87, 97	H057, H067, H077, H087, H097	S, N, N, N, C

Probabilidad de inclusión de cada hogar: 1/k = 1/10 = 0,10.

Paso 3. Comprobación e interpretación. a) Sin repetidos y dentro del marco: 10 identificadores distintos, del 7 al 97. b) Periodicidad: Adrián revisa el orden de la lista; las altas están por fecha, entre una y seis al día, sin bloques fijos, y la zona de alta no se repite cada 10 posiciones. c) Composición: la selección tiene 4 hogares del Norte, 4 del Centro y 2 del Sur (el marco, 40, 36 y 24). Que se parezca no prueba nada por sí solo, pero un resultado muy distinto obligaría a revisar el orden de la lista. d) Qué no permite: no es un aleatorio simple (solo había 10 muestras posibles); con 10 hogares, cualquier porcentaje será muy impreciso; y nada se puede decir de los hogares que no están en la lista.

RESULTADO

Selección sistemática con arranque aleatorio 7 e intervalo 10: **H007, H017, H027, H037, H047, H057, H067, H077, H087 y H097**, cada hogar con probabilidad 0,10; el marco no tiene periodicidad que coincida con k. Conclusiones válidas solo para los 100 hogares de la lista de interés, con el arranque, la versión del marco y el registro de contactos guardados en el protocolo.

Comprueba que lo has entendido. a) Si el arranque hubiera sido 3, ¿cuáles serían el primer y el último hogar? b) ¿Por qué no se puede empezar siempre por el 1? c) ¿Qué cambiaría si la lista estuviera ordenada por portales de diez viviendas?

PRÁCTICA

Las llamadas de revisión de Ruedalia

Caso ficticio. Txema quiere saber qué horario de revisión prefieren los clientes particulares con cuota de mantenimiento. Ousmane tiene la lista: 60 clientes numerados del 1 al 60 por fecha de alta, sin duplicados. Solo hay tiempo para llamar a 6. Txema propone «llamar a uno de cada diez, empezando por el que salga en el sorteo»; en el sorteo entre 1 y 10 sale el 4.

Encargo. 1) Define población, marco, unidad de muestreo e informante. 2) Calcula el intervalo y escribe la selección con sus identificadores. 3) Comprueba repetidos, límites del marco y periodicidad. 4) Di qué tipo de muestreo es y qué no es. 5) Escribe la decisión y sus limitaciones.

Plantilla IC-P07 · Selección sistemática de Ruedalia

Campo	Respuesta
Datos y fuentes con unidad y periodo	
Procedimiento o cálculo	
Resultado y evidencia	
Decisión y limitaciones	

Lista de cotejo

☐ Población, marco y unidad definidos.
☐ k = N/n con su unidad.
☐ Seis identificadores dentro del marco.
☐ Periodicidad revisada.
☐ Tipo de muestreo bien nombrado.
☐ Límite de la conclusión escrito.

LECTURA Y DEBATE

Treinta respuestas en la puerta



Relato ficticio. El viernes 2 de abril, antes de que nadie escribiera un protocolo, Nora pidió a Adrián que se pusiera en la puerta de la tienda con la tableta. De diez a doce, Adrián paró a treinta personas y les hizo una sola pregunta: «¿Compraría un kit de comida preparada para casa?». Veinticuatro dijeron que sí.

El lunes, Nora entró en la trastienda con el resultado en la mano. «Ochenta por ciento. Ya está: el barrio quiere kits. Encargamos envases para el doble».

Adrián, que lo había apuntado todo, abrió su hoja. Junto a cada respuesta había anotado si la persona tenía tarjeta de cliente: veintiséis de las treinta la tenían. Ninguna venía de una oficina: a esa hora están trabajando. Olvido, desde el obrador, añadió lo que ella sabía sin hoja: «Por las mañanas vienen las de siempre. Los que quiero ver son los que se van al supermercado».

El dato no era falso: veinticuatro de treinta habían dicho que sí. Lo que Nora no sabía era a quién representaban. Aquella tarde, el equipo de análisis le propuso tratar el sondeo como lo que era, una primera lectura de clientes habituales, y no encargar envases hasta tener la selección del marco depurado. Nora aceptó, con una condición: «El viernes quiero el protocolo por escrito».

Preguntas. 1) ¿Qué afirmación del caso necesita una comprobación antes de convertirse en decisión? 2) ¿Qué información pidió primero el equipo y por qué cambió la respuesta de Nora? 3) Propón una actuación provisional que permita aprender sin ocultar la incertidumbre.

APRENDE CON LA IA

Unai prepara las 20 entrevistas con cuotas y pregunta a un asistente de IA si con ellas puede dar un margen de error en el informe. Esta es la respuesta:

RESPUESTA DE LA IA

«Sí. Cumplir cuotas de edad convierte automáticamente la muestra en probabilística, porque la composición de la muestra coincide con la de la población. Con 20 entrevistas puedes calcular el margen de error con la fórmula del aleatorio simple y decir que el resultado vale para todo el barrio. Si alguna persona no quiere contestar, sustitúyela por la siguiente que pase, así mantienes las cuotas.»

Tu tarea. Localiza los errores de la respuesta, contrasta cada uno con lo que has visto en 7.2, 7.4 y 7.5 y escribe una respuesta corregida que diga qué se puede afirmar con esas 20 entrevistas y cuál es su límite. Conserva la prueba de cada corrección (definición, tabla o ejemplo de la unidad).

Ojo con la IA. Puede aplicar fórmulas correctas a supuestos que no se cumplen: pregunta qué condiciones exige y compruébalas tú.

ACTIVIDADES FINALES

Comprueba

14. Resume una distinción decisiva de esta unidad (probabilístico y no probabilístico, estrato y conglomerado, o error de muestreo y sesgo) y aplícala al caso resuelto.
15. Explica qué diferencia hay entre la población objetivo y la población de estudio en el Estudio de kits, y cuál de las dos describe la muestra del caso.
16. ¿Por qué el sistemático del caso tiene la misma probabilidad de inclusión que un aleatorio simple y, aun así, no es un aleatorio simple?

Aplica

17. Repite la práctica de Ruedalia cambiando una condición (la lista ordenada por día de revisión, con ciclos de cinco días laborables, o un marco de 63 clientes con k = 10 y recorrido circular). Señala qué resultado cambia y cuál se mantiene.
18. La lista de interés se amplía a 120 hogares y Nora quiere 12 entrevistas. Calcula el intervalo, escribe la selección con arranque 5 y di qué parte del protocolo hay que actualizar.
19. Diseña un muestreo en dos etapas para las oficinas del polígono: primera etapa, empresas del directorio; segunda, personas que comen en la empresa un día sorteado. Escribe el marco y el sorteo de cada etapa.

Investiga

20. Localiza una fuente primaria o documentación técnica sobre sesgos de selección y selección documentada (la metodología de una encuesta del INE, un código profesional de investigación de mercados o un manual universitario de muestreo). Registra la fecha de consulta, el ámbito, la afirmación que sustenta y una limitación.
21. Busca en la nota de una encuesta oficial cómo describe su diseño muestral (tipo de muestreo, etapas, tamaño) y explica con tus palabras qué se sortea en cada etapa.

TU INVESTIGACIÓN AVANZA · PASO 7 · I07 · PROTOCOLO DE SELECCIÓN

Para el negocio de tu investigación, con datos ficticios o propios autorizados: 1) población objetivo y población de estudio; 2) marco con versión, fecha y registro de depuración (filas, duplicados, ajenos); 3) unidad de muestreo, de análisis e informante; 4) método elegido y alternativa descartada, con su motivo; 5) selección hecha con su arranque o semilla y archivo guardado; 6) reglas de sustitución y registro de contactos; 7) responsable y fecha de revisión. Versiona el protocolo: si cambias algo, conserva el motivo.

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

☐ Hecho ☐ Revisado ☐ Guardado

OJO PROFESIONAL

Documento: borrador del protocolo de selección EK-2027/PS-v0.1

Documento ficticio, preparado por Unai el miércoles 07/04/2027 para que lo revise el equipo de análisis. Contiene **cinco errores**.

N.º	Campo	Contenido
1	Población de estudio	Hogares de la lista de interés de kits (11/01 - 31/03/2027)
2	Marco	Lista de interés original, 104 filas, tal como se exportó de la web y de la tienda
3	Método	Muestreo aleatorio simple: 20 entrevistas con cuotas de edad (10 de menos de 45 años y 10 de 45 o más) en la puerta de la tienda
4	Sistemático (alternativa)	Intervalo k = 10; empezamos por el primero de la lista: H001, H011, H021…
5	Probabilidad	Cada hogar de la lista tiene una probabilidad de 0,10
6	Sustituciones	Si alguien no contesta, se llama al siguiente de la lista
7	Conclusión prevista	Los resultados se presentarán como opinión del barrio, con su margen de error
8	Responsable	Diseño: equipo de análisis · registro: Adrián · aprueba: Nora

Encargo. Señala los cinco errores, explica por qué lo son según la unidad y reescribe cada línea afectada para que el protocolo se pueda usar. **Para pensar:** ¿qué línea es correcta solo si se corrigen antes otras dos?

N.º	Línea	Por qué es un error	Corrección
1			
2			
3			
4			
5			

AUTOEVALUACIÓN

1. 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.

2. ¿Cuál permite calcular un margen de error?
a) 20 entrevistas con cuotas de edad.
b) 10 hogares sorteados sin reposición de un marco numerado.
c) 286 respuestas de una encuesta abierta en la web.

3. Marco de 100 hogares, n = 10, arranque 7. La selección sistemática es...
a) 7, 14, 21... 70.
b) 1, 11, 21... 91.
c) 7, 17, 27... 97.

4. Unai llena en la puerta cuotas de edad. El grupo peor representado será...
a) el de quienes no pasan por la tienda.
b) el de menos de 45 años, porque la cuota es de 10.
c) ninguno, porque las cuotas lo impiden.

5. Oficinas pequeñas sin directorio fiable: para estimar con margen de error conviene...
a) una bola de nieve desde Albada Asesores.
b) un muestreo por áreas que registre y sortee las oficinas de cada zona elegida.
c) preguntar a las oficinas clientes.

6. En el caso resuelto, cada hogar tiene una probabilidad de inclusión de...
a) 0,10.
b) 0,07.
c) 0,01.

7. Ruedalia: 60 clientes, 6 llamadas, arranque 4. Es correcto que...
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.

8. «Cumplir cuotas de edad convierte automáticamente la muestra en probabilística» es falso porque...
a) las cuotas exigen más de 100 entrevistas.
b) dentro de cada cuota elige el entrevistador y la probabilidad de inclusión no se conoce.
c) la edad no sirve como cuota.

9. En el Ojo profesional, «empezamos por el primero de la lista» se corrige...
a) sorteando el arranque entre 1 y k.
b) empezando por el último.
c) cambiando k por 11.

10. En tu I07, ¿qué dato te haría revisar la selección antes del campo?
a) Filas duplicadas o lista ordenada por portales.
b) La fecha en el nombre del archivo.
c) Un arranque impar.

Ya sé...

- ☐ definir población objetivo, población de estudio, marco y unidad, y depurar el marco.
- ☐ seleccionar un aleatorio simple sin reposición y un sistemático con arranque aleatorio de forma reproducible.
- ☐ distinguir estratos, conglomerados, áreas, etapas y rutas, y elegir el esquema adecuado.
- ☐ explicar qué permiten concluir la conveniencia, el juicio, las cuotas y la bola de nieve.
- ☐ detectar una conclusión no sustentada por la selección y entregar un protocolo revisable.



UNIDAD 8

Tamaño, precisión y no respuesta

Bloque 3 · Muestreo y preparación de datos

EN ESTA UNIDAD APRENDERÁS A…

- Explicar cómo influyen la confianza, el error, la variabilidad, la población y el diseño en el tamaño de la muestra, y calcular el error de muestreo de un tamaño dado.
- Aplicar la corrección para población finita y saber cuándo basta la fórmula de población grande.
- Calcular el tamaño, redondearlo y distinguir respuestas completas de personas que hay que contactar.
- Repartir la muestra entre estratos y comprobar las sumas y sus consecuencias.
- Planificar los contactos con una tasa de respuesta prevista y explicar lo que el tamaño no corrige.

SITUACIÓN DE PARTIDA

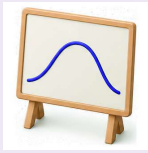
Lunes 12 de abril de 2027. Quique, comercial de una empresa de encuestas en línea, enseña a Nora su oferta: «Quinientas encuestas en una semana, con un error del ±4,4 % al 95 % de confianza». Nora está tentada: «Son muchas más de las que podemos hacer nosotros por teléfono».

Tu equipo de análisis tiene hasta el viernes 23 para entregar el tamaño y el plan de contactos del Estudio de kits:

- cuántas respuestas completas hacen falta para la precisión que Nora necesita, y con qué supuestos;
- a cuántos hogares hay que llamar para conseguirlas y cómo se reparten por zonas;
- qué garantiza y qué no garantiza un «±4,4 %».

PARA EMPEZAR

1. ¿Qué debe decidir o comprobar Nora antes de comparar la oferta de Quique con un estudio propio?
2. ¿Qué sabemos de esas 500 encuestas y qué estamos suponiendo?
3. Si responde el 8 % de los invitados, ¿cuántas respuestas dan 500 invitaciones? ¿Y 6.250?
4. ¿Qué prueba permitiría revisar la primera respuesta de Nora?



Confianza y error 8.1



Población finita 8.2



Cálculo y redondeo 8.3



Estratos 8.4



No respuesta 8.5

8.1 Confianza y error

Cuántas respuestas hacen falta depende de cuánta precisión necesita la decisión, con qué seguridad y sobre qué población. El número no sale de la costumbre ni del presupuesto.

Las cinco variables del tamaño. Para estimar una proporción (qué parte de los hogares compraría un kit), el tamaño de una muestra probabilística depende de:

- **Error de muestreo admitido (e)**: la precisión, en tanto por uno. Con e = 0,05, el resultado se da con ±5 puntos porcentuales. Cuanto más pequeño, más muestra.
- **Nivel de confianza**: la proporción de veces que acertaría el procedimiento si se repitiera muchas veces con muestras distintas. Se traduce en un coeficiente **z** de la distribución normal.
- **Variabilidad (p · q)**: cuánto se espera que se repartan las respuestas; p es la proporción esperada y q = 1 − p.
- **Tamaño de la población (N)**: pesa poco salvo que la muestra sea una parte apreciable de ella (8.2).
- **Diseño**: las fórmulas de esta unidad son las del **aleatorio simple sin reposición** (unidad 7). Un estratificado bien hecho puede necesitar algo menos; un muestreo por conglomerados, más.

Tabla 1 · Niveles de confianza y coeficiente z

Nivel de confianza	90 %	95 %	99 %
z	1,645	1,96	2,576

Cálculo del error de muestreo dado n. Si el tamaño ya está fijado, se despeja el error. En un aleatorio simple de una población grande:

$$e = z \cdot \sqrt{(p \cdot q / n)}$$

Es la fórmula que usa Quique. Su oferta: n = 500, 95 % (z = 1,96), p = q = 0,5. Entonces p · q / n = 0,25 / 500 = 0,0005; √0,0005 = 0,02236; e = 1,96 · 0,02236 = **0,0438**, es decir, **±4,4 puntos**. El cálculo es correcto, pero **solo si las 500 respuestas fueran un aleatorio simple** de la población que interesa. Con 200 respuestas, el error sería de ±6,9 puntos; con 100, de ±9,8: para reducir el error a la mitad hay que multiplicar el tamaño por cuatro. Con este cálculo puedes revisar la ficha técnica de cualquier encuesta publicada.

Tabla 2 · Relación entre error, confianza y tamaño

(población grande, p = q = 0,5; respuestas completas, redondeadas al alza)

Error admitido (puntos)	90 %	95 %	99 %
±3	752	1.068	1.844
±4	423	601	1.037
±5	271	385	664

Cómo se lee el nivel de confianza. «Al 95 %» no significa que el 95 % de los hogares piense lo mismo ni que el resultado acierte con probabilidad 0,95. Significa que, si se repitiera el procedimiento con muchas muestras aleatorias del mismo marco, en unas 95 de cada 100 el resultado quedaría a menos de e del valor verdadero. Es una propiedad **del procedimiento**; por eso exige que la selección sea aleatoria.

En la hoja de cálculo. El error de la oferta se calcula con =1,96*RAIZ(0,25/500). El coeficiente z sale de la normal: =DISTR.NORM.ESTAND.INV(0,975) devuelve 1,96 para el 95 %; con 0,95, 1,645 (90 %), y con 0,995, 2,576 (99 %). Escribe los supuestos en celdas propias.

Errores de muestreo y errores ajenos al muestreo. El error de la fórmula mide solo la variación debida a observar una muestra y no a toda la población. **No incluye** los sesgos de cobertura, selección, no respuesta y medición, que no bajan al aumentar la muestra (unidad 7).

CLAVE · «±4,4 %» ES UN RESULTADO CONDICIONADO

Un margen de error solo vale si la muestra es probabilística y procede del marco de la población estudiada. Pregunta siempre cómo se seleccionó y cuántos contestaron.

ACTIVIDADES

1. Explica, con una frase cada uno, el efecto en el tamaño de la confianza, el error, la variabilidad, la población y el diseño. Después realiza dos cambios de parámetro con la tabla 2 (por ejemplo, de 95 % a 99 % con ±5 %, y de ±5 % a ±4 % con el 95 %) y di cuánto cambia el tamaño.
2. Calcula el error de muestreo, con el 95 % y p = q = 0,5 (población grande), para 500, 200 y 100 respuestas. Repite el de 500 con p = 0,2 y explica por qué baja.

8.2 Variabilidad y población finita

Cuando la muestra es una parte apreciable de la población, cada respuesta aporta más información: la fórmula lo recoge con la corrección para población finita.

La variabilidad. El producto p · q es máximo cuando p = 0,5 y baja al alejarse de ese valor: con p = 0,1, p · q = 0,09; con 0,2, 0,16; con 0,3, 0,21; con 0,4, 0,24; y con 0,5, 0,25. Si no hay un dato fiable de un estudio anterior, se toma **p = q = 0,5**, el supuesto más prudente: da el tamaño mayor.

De dónde sale la corrección. En un aleatorio simple **sin reposición**, la variabilidad de la proporción muestral es p · q / n multiplicada por el **factor de corrección para población finita** (N − n)/(N − 1): cuanto mayor es la parte de la población que entra en la muestra, menos varía el resultado. Se exige que z veces la desviación sea igual al error admitido y se despeja n:

1. $e = z \cdot \sqrt{[(p \cdot q / n) \cdot (N - n)/(N - 1)]}$.
2. Al cuadrado: $e^2 \cdot n \cdot (N - 1) = z^2 \cdot p \cdot q \cdot (N - n)$.

1. Se agrupan los términos con n: $n \cdot [e^2 \cdot (N - 1) + z^2 \cdot p \cdot q] = N \cdot z^2 \cdot p \cdot q$.
2. Resultado: $\mathbf{n = N \cdot z^2 \cdot p \cdot q \,/\, [e^2 \cdot (N - 1) + z^2 \cdot p \cdot q]}$.

Cuánto pesa el factor. Con el listado de contactos del estudio (N = 1.000) y n = 278, (N − n)/(N − 1) = 722/999 = 0,7227 y su raíz, 0,850: el error baja un 15 % frente al de N grande.

Población grande. Si N es muy grande frente a n, el factor vale casi 1 y queda la fórmula de la población grande (o infinita): $\mathbf{n_o = z^2 \cdot p \cdot q \,/\, e^2}$. Con el 95 %, ±5 % y p = q = 0,5: $n_o = 3{,}8416 \cdot 0{,}25 / 0{,}0025 = 384{,}16 \rightarrow 385$. Las dos fórmulas están relacionadas: $n = n_o / [1 + (n_o - 1)/N]$. Para el marco del Estudio de kits, N = 1.000: 384,16 / 1,38316 = 277,74.

Tabla 3 · El tamaño según la población

(95 %, ±5 %, p = q = 0,5, aleatorio simple sin reposición)

Población N	300	1.000	3.000	6.000	10.000	100.000	Grande
n exacto	168,70	277,74	340,65	361,10	369,98	382,69	384,16
n redondeado	169	278	341	362	370	383	385

A partir de unos miles de unidades, la población casi no cambia el tamaño: un barrio de unos 6.000 hogares necesita 362 respuestas y una ciudad de 100.000, 383. Lo que pesa es la precisión que se pide.

¿Cuándo basta la población grande? Una regla práctica: si n_o es menos del 5 % de N, la corrección apenas cambia el resultado. Con N = 10.000, 385 es el 3,85 %: la fórmula de población grande da 385 y la finita 370. Con N = 1.000, 385 sería el 38,5 %: hay que corregir.

EN TU SECTOR

Poblaciones pequeñas: a veces, censo. Las empresas del directorio del polígono son 30. Con el 95 %, ±5 % y p = 0,5, n = 30 · 0,9604 / (0,0025 · 29 + 0,9604) = 27,89 → 28: casi todas. En poblaciones pequeñas de empresas o tiendas se suele preguntar a todas (censo) y trabajar la no respuesta.

Hipótesis a la vista. Las fórmulas suponen: 1) aleatorio simple sin reposición en un marco que coincide con la población de estudio; 2) estimación de una proporción; 3) que todas las unidades seleccionadas responden. Si alguna no se cumple, el número sigue saliendo, pero ya no significa lo mismo.

ACTIVIDADES

3. Aplica 8.2 al conflicto de la unidad con la ficha de I08: documenta con un ejemplo qué población describiría el estudio propio (marco de 1.000 contactos) y cuál la oferta de Quique, propón una alternativa (por ejemplo, usar sus encuestas solo como exploración) y escribe la condición que te haría cambiar de elección.
4. Calcula el tamaño para N = 6.000 con la fórmula finita y con n_o y la relación entre ambas. Explica por qué la diferencia con N = 1.000 es de solo 84 respuestas.

8.3 Cálculo y redondeo

Una fórmula bien aplicada exige escribir los supuestos, sustituir con cuidado, redondear siempre hacia arriba y no confundir a quien responde con a quien se llama.

El procedimiento en cinco pasos

1. Escribir los supuestos: población y marco (N), diseño, confianza (z), error (e) y p.
2. Sustituir y calcular por separado el numerador y el denominador, con cuatro decimales.
3. Dividir y **redondear al entero superior**, aunque el decimal sea pequeño: con una respuesta menos, el error supera el admitido.
4. Comprobar: calcular el error con el tamaño redondeado, que no debe superar el admitido.
5. Pasar de **respuestas completas** a **contactos**: contactos = n redondeado ÷ tasa de respuesta prevista, también al alza.

Respuestas completas y contactos. El tamaño n es el número de cuestionarios **completos y válidos** que hacen falta. Como no todos los hogares seleccionados contestan, hay que seleccionar y contactar a más. La **tasa de respuesta prevista** se toma de un estudio anterior o de una prueba; si no la hay, se declara como supuesto y se revisa en las primeras llamadas. En el Estudio de kits sale de una prueba de llamadas de marzo: contestaron 21 de 30 hogares (70,0 %).

Ejemplo. Con N = 1.000, el 95 %, ±4 % y p = q = 0,5: n = 960,4 / (0,0016 · 999 + 0,9604) = 960,4 / 2,5588 = 375,33 → **376** respuestas completas. Con una respuesta prevista del 70 %: 376 / 0,70 = 537,14 → **538 contactos** (537 · 0,70 = 375,9 no llega a 376). Comprobación: con 376, e = 0,0399, por debajo de 0,04.

En la hoja de cálculo. Con los supuestos en celdas con nombre: =REDONDEAR.MAS(N*z^2*p*q/(e^2*(N-1)+z^2*p*q);0) para las respuestas y =REDONDEAR.MAS(n/tasa;0) para los contactos (en inglés, ROUNDUP). Comprueba el resultado a mano una vez.

Errores frecuentes. Escribir el error en porcentaje (5 en lugar de 0,05); usar el z del 95 % para el 99 %; redondear al entero más próximo; calcular los contactos desde 277,74 o dividir por 1,70 en lugar de 0,70; olvidar que los hogares que resulten fuera del marco no cuentan como respuestas; y comparar tamaños calculados con supuestos distintos sin decirlo (otra confianza, otro p, otro diseño).

PARA SABER MÁS

El efecto del diseño. Si no es un aleatorio simple, el tamaño se multiplica por el efecto del diseño, que se estima con estudios anteriores. Con conglomerados de manzanas y un efecto de 1,5, las 278 respuestas pasarían a 278 · 1,5 ≈ 417 (ajuste conservador). En un estratificado proporcional suele ser algo menor que 1.

Cuadro 1 · Cambios de parámetro con N = 1.000

(aleatorio simple sin reposición)

Supuesto que cambia	Resto igual (95 %, ±5 puntos, p = 0,5)	Respuestas completas
Ninguno (caso base)	—	278
Confianza 99 %	z = 2,576	400
Confianza 90 %	z = 1,645	214
Error ±4 puntos	e = 0,04	376
p = 0,3 (de un estudio previo)	p · q = 0,21	245
Población grande (sin factor)	—	385



ACTIVIDADES

5. Calcula el tamaño para un listado ficticio de 2.400 hogares de otra zona de reparto (95 %, ±5 %, p = q = 0,5): redondea, asígnalo a sus tres estratos (1.200, 720 y 480 hogares) y planifica los contactos con una respuesta prevista del 60 %.
6. Adrián redondea 375,33 a 375 «porque el decimal es pequeño». Calcula el error con 375 y con 376 y explica qué regla del procedimiento se ha saltado.

8.4 Asignación por estratos

Si la muestra se estratifica, hay que decidir cuántas respuestas van a cada estrato. Cada reparto tiene consecuencias para la precisión y para cómo se calculan los resultados.

Afijación proporcional. Cada estrato recibe una parte de la muestra igual a su peso en la población: n_h = n · N_h / N. Las cuotas exactas casi nunca son enteras: se toma la parte entera de cada una y las unidades que faltan se dan a los estratos con **mayores restos**. Después se **comprueba que la suma** es n. La muestra queda **casi autoponderada**: cada respuesta representa a unos 3,6 hogares del marco. Es el reparto por defecto cuando lo que interesa es, sobre todo, el resultado del total.

Tabla 4 · Afijación proporcional del Estudio de kits

(marco de 1.000 contactos; n = 278; respuesta prevista 70 %)

Zona	Hogares N_h	Peso	Cuota exacta	Respuestas	Contactos
Norte	420	42,0 %	116,76	117	168
Centro	350	35,0 %	97,30	97	139
Sur	230	23,0 %	63,94	64	92
Total	**1.000**	**100,0 %**	**278,00**	**278**	**399**

Las partes enteras suman 116 + 97 + 63 = 276; faltan 2, que van al Sur (resto 0,94) y al Norte (0,76). Los contactos se calculan en cada estrato y al alza: 117 / 0,70 = 167,14 → 168; 97 / 0,70 = 138,57 → 139; 64 / 0,70 = 91,43 → 92. Suman **399**, uno más que los 398 del total, porque cada estrato redondea por su lado: se planifican 399 para garantizar los 117, 97 y 64. Después se sortean los hogares dentro de cada zona con un aleatorio simple (unidad 7), con la semilla anotada, y se sortea también una reserva por zona.

Alternativas. En la **afijación igual**, cada estrato recibe lo mismo (93, 93 y 92). El Sur gana precisión (con la corrección finita y N = 230, 64 respuestas dan ±10,4 puntos y 92, ±7,9), pero la muestra deja de ser autoponderada: cada respuesta del Norte representa 420/93 = 4,52 hogares; una del Sur, 230/92 = 2,50. El total se calcula **ponderando** cada estrato por su peso, y el error del conjunto se recalcula. En la **afijación óptima** (de Neyman) se da más muestra a los estratos más variables o más baratos de encuestar; exige conocer esa variabilidad. La afijación igual conviene cuando comparar zonas importa tanto como el total.

Por qué importa ponderar. Supón que, con afijación igual, las tasas hipotéticas de sí al kit son el 50 % del Norte, el 40 % del Centro y el 30 % del Sur. La media simple es (93 · 0,50 + 93 · 0,40 + 92 · 0,30) / 278 = 111,3 / 278 = 40,0 %, porque el Sur pesa en la muestra más que en el marco. Ponderando por zona: 0,42 · 50 + 0,35 · 40 + 0,23 · 30 = **41,9 %**. Con afijación proporcional, las dos cuentas casi coinciden.

CLAVE · CADA ESTRATO TIENE SU PROPIO ERROR

El ±5 % es del total. Si Nora quiere conclusiones por zona, el error de cada zona es mayor: decide antes si se necesitan y cuántas respuestas pide cada una.

ACTIVIDADES

7. Comprueba la tabla 4: calcula las cuotas, aplica los mayores restos y los contactos de cada zona. Explica por qué salen 399 contactos y no 398, y qué harías si solo hubiera presupuesto para 398.

8. Aplica 8.4 al conflicto de la unidad con la ficha de I08: Nora quiere comparar el Sur con el resto. Documenta la afijación proporcional y una alternativa (igual u otra), con su efecto en la ponderación y en el error de cada zona, y escribe la condición que te haría elegir una u otra.

8.5 No respuesta y limitaciones

Llamar a más hogares asegura el número de respuestas, no que quienes responden se parezcan a quienes no lo hacen.

El embudo de respuesta. Entre los hogares seleccionados y los cuestionarios completos hay pérdidas: no localizados (no cogen el teléfono tras los intentos previstos), rechazos e incompletos. Cada contacto se anota en el **registro de contactos** (fecha, hora, intento y resultado) y al final se calculan las tasas. La **tasa de respuesta** es completas ÷ hogares seleccionados del marco (se descuentan los que resulten estar fuera del marco).

Tipos de no respuesta. La **no respuesta total** es la del hogar que no contesta nada (no localizado o rechazo). La **no respuesta parcial** es la de quien contesta pero deja preguntas en blanco: se trata en la depuración (unidad 9). La **tasa de cooperación** (aceptan ÷ localizados) ayuda a separar el problema de horario (no se encuentra a nadie) del de interés (se encuentra y no quiere). Las dos piden remedios distintos.

Tabla 5 · Resultado posible del campo

(simulación ficticia del equipo para preparar el plan)

Zona	Contactos	Completas	Tasa de respuesta	Peso entre las respuestas	Peso en el marco
Norte	168	101	60,1 %	36,3 %	42,0 %
Centro	139	104	74,8 %	37,4 %	35,0 %
Sur	92	73	79,3 %	26,3 %	23,0 %
Total	**399**	**278**	**69,7 %**	**100,0 %**	**100,0 %**

Tabla 6 · Embudo de respuesta del total

(misma simulación)

Etapa	Hogares	Sobre la etapa anterior	Pérdida
Seleccionados y contactados	399	—	—
Localizados (máximo 3 intentos)	336	84,2 %	63 no localizados
Aceptan la entrevista	284	84,5 %	52 rechazos
Cuestionarios completos	278	97,9 %	6 incompletos

Qué dice la simulación. Se llega a 278 completas, pero **no** con el reparto previsto: al Norte le faltan 16 respuestas (101 de 117) y el Sur tiene 9 de más. Si los hogares del Norte que no contestan son distintos (por ejemplo, trabajan hasta tarde y pedirían más reparto), el resultado tendrá un **sesgo de no respuesta** que ningún aumento de llamadas corrige por sí solo. La tasa del Norte está 9,6 puntos por debajo de la del total: si se hubiera planificado con ella, el Norte habría necesitado 117 / 0,601 = 194,7 → 195 contactos, no 168.

Qué se puede hacer

1. **Reintentar** en horarios distintos antes de dar por perdido un hogar: es lo que más reduce la no respuesta.
2. **Ampliar contactos** en el estrato que falla con la reserva sorteada: para 16 respuestas al 60,1 %, 16 / 0,601 = 26,6 → unos 27 contactos más.
3. **Comparar perfiles**: zona, canal y antigüedad de quienes responden frente al marco.
4. **Ponderar** por zona para que cada una pese lo que pesa en el marco; corrige la composición por zona, no las diferencias dentro de cada zona.
5. **Informar**: tasa de respuesta, perfiles y limitaciones van en el informe.

Lo que no se hace: sustituir a quien no contesta por el vecino o por un voluntario, ni subir el tamaño «por si acaso» sin mirar quién falta.

LA NORMA DICE

Ley 11/2022, de 28 de junio, General de Telecomunicaciones, art. 66.1 b): los usuarios finales tienen derecho «a no recibir llamadas no deseadas con fines de comunicación comercial, salvo que exista consentimiento previo del propio usuario para recibir este tipo de comunicaciones comerciales o salvo que la comunicación pueda ampararse en otra base de legitimación de las previstas en el artículo 6.1 del Reglamento (UE) 2016/679». Una llamada de estudio no puede usarse para vender kits (finalidad, RGPD 5.1 b). Consulta la versión vigente.



ACTIVIDADES

9. Con la tabla 5, calcula las tasas de respuesta y los pesos de cada zona y explica por qué llegar a 278 completas no basta para dar por bueno el campo.
10. Aplica 8.5 al conflicto de la unidad: la oferta de Quique invitó a 6.250 personas de su panel para conseguir 500 respuestas. Calcula la tasa de participación, documenta qué no sabes de quienes no contestaron, propón una alternativa y escribe la condición que te haría aceptar sus datos.
11. Diseña la hoja del registro de contactos del Estudio de kits (columnas, códigos de resultado y número máximo de intentos) y la fórmula de la tasa de respuesta por zona.

CASO RESUELTO · PASO A PASO

Tamaño y contactos del Estudio de kits

Caso ficticio. El viernes 23 de abril de 2027 el equipo entrega a Nora el tamaño de la encuesta telefónica sobre la entrega de kits.

Paso 1. Decisión, datos y alcance. Decisión: cuántas respuestas completas hacen falta para estimar la proporción de hogares interesados con ±5 puntos al 95 % de confianza, y a cuántos hogares hay que llamar. Datos: marco = listado de contactos autorizados 2027 (LC-v1.0, más amplio que el marco de 100 hogares de la UD7), N = **1.000 contactos** (hogares) de la zona de reparto que compraron y aceptaron llamadas para estudios; z = 1,96; p = q = 0,5 (sin dato previo); e = 0,05; aleatorio simple sin reposición. Respuesta prevista: 70 %. Alcance: los 1.000 contactos del listado, no los unos 6.000 hogares del barrio.

Paso 2. Procedimiento. Numerador: $N \cdot z^2 \cdot p \cdot q = 1.000 \cdot 3{,}8416 \cdot 0{,}25 = 960{,}4$. Denominador: $e^2 \cdot (N - 1) + z^2 \cdot p \cdot q = 0{,}0025 \cdot 999 + 0{,}9604 = 2{,}4975 + 0{,}9604 = 3{,}4579$.

$n = 960{,}4 / 3{,}4579 = 277{,}74 \rightarrow$ **278 respuestas completas** (al entero superior).

Contactos: $278 / 0{,}70 = 397{,}14 \rightarrow$ **398 contactos** ($397 \cdot 0{,}70 = 277{,}9$ no llega; $398 \cdot 0{,}70 = 278{,}6$ sí).

Paso 3. Comprobación e interpretación. a) Error con el tamaño redondeado: con 278, $e = 1{,}96 \cdot \sqrt{(0{,}25/278)} \cdot \sqrt{(722/999)} = 0{,}0500$ (no supera 0,05); con 277 sería 0,0501, por encima del admitido. b) La corrección para población finita ahorra 107 respuestas frente a la fórmula de población grande (385). c) Reparto por zonas (8.4): 117, 97 y 64, que suman 278; reparto estratificado proporcional, de error no superior al del aleatorio simple (±5 es cota). d) Qué no permite: el ±5 % solo mide el error de muestreo de un aleatorio simple del listado; no corrige la cobertura (hogares sin autorización ni compra) ni la no respuesta. e) Plan de campo: sorteo de los hogares de cada zona con su reserva, registro de cada intento y revisión de las tasas por zona tras 100 llamadas, para ampliar contactos donde falten.

RESULTADO

Con N = 1.000, el 95 % de confianza, p = q = 0,5 y e = 0,05, n = 277,74 → **278 respuestas completas**; con una respuesta prevista del 70 %, **contactar al menos a 398 hogares**. El plan registra cada contacto y compara los perfiles de quienes responden con el marco: más contactos no eliminan el sesgo de no respuesta.

Comprueba que lo has entendido. a) ¿Cuántas respuestas harían falta con ±4 %? b) Si la respuesta prevista fuera del 50 %, ¿cuántos contactos? c) ¿Por qué no se usa N = 6.000?

PRÁCTICA

La encuesta de satisfacción de Ruedalia

Caso ficticio. Txema quiere saber qué proporción de sus clientes está satisfecha con el aviso de revisión por mensaje. Ousmane exporta la base de clientes: 500 clientes con teléfono y autorización de contacto, sin duplicados. Txema pide el mismo rigor que el Estudio de kits: 95 % de confianza, ±5 puntos, sin dato previo de la proporción y aleatorio simple sin reposición. En una prueba de llamadas, 7 de cada 10 clientes contestaron.

Encargo. 1) Escribe los supuestos y los datos con su unidad. 2) Calcula el tamaño con la fórmula de población finita y redondea. 3) Calcula los contactos con la respuesta prevista. 4) Comprueba el error con el tamaño redondeado. 5) Escribe la decisión y sus limitaciones.

Plantilla IC-P08 · Tamaño y contactos de Ruedalia

Campo	Respuesta
Datos y fuentes con unidad y periodo	
Procedimiento o cálculo	
Resultado y evidencia	
Decisión y limitaciones	

Lista de cotejo

☐ Supuestos escritos (N, z, p, e, diseño).
☐ Numerador y denominador por separado.
☐ Tamaño al entero superior.
☐ Contactos desde el tamaño redondeado y al alza.
☐ Error comprobado.
☐ Límite de la conclusión escrito.

LECTURA Y DEBATE

Quinientas respuestas en una semana



Relato ficticio. Quique dejó sobre el mostrador una carpeta con gráficos de colores. «Nuestro panel tiene miles de personas de la provincia. En una semana te devuelvo quinientas encuestas completas, con un error del ±4,4 % al 95 %. Tu equipo, por teléfono, no llega ni a trescientas».

Nora miró la carpeta y después a Adrián, que estaba ordenando el registro de la unidad anterior. «Quinientas son más que doscientas setenta y ocho, ¿no? Pues mejor».

Adrián, antes de opinar, hizo lo que hacía siempre: preguntar por los números que faltaban. «¿A cuántas personas les enviáis la invitación para conseguir esas quinientas?». Quique buscó en su tableta: «Normalmente, a unas seis mil doscientas cincuenta. Contesta el que quiere; por eso es tan rápido». «¿Y cuántas viven en nuestra zona de reparto?». Quique se encogió de hombros: «No lo filtramos por barrio, pero se puede».

Esa tarde el equipo de análisis hizo la cuenta para Nora: 500 de 6.250 es una participación del 8,0 %. El ±4,4 % era la fórmula del aleatorio simple aplicada a personas que se apuntan a un panel y eligen responder, de toda la provincia. Nadie sabía en qué se parecían a los clientes de La Despensa. Nora no firmó. «Me quedo con nuestras 278, pero quiero ver cuántas llamadas hacen falta y quién no contesta».

Preguntas. 1) ¿Qué afirmación del caso necesita una comprobación antes de convertirse en decisión? 2) ¿Qué información pedirías primero y por qué cambiaría la respuesta? 3) Propón una actuación provisional que permita aprender sin ocultar la incertidumbre.

APRENDE CON LA IA

Unai pregunta a un asistente de IA si pueden ahorrarse las llamadas publicando la encuesta en las redes de La Despensa. Esta es la respuesta:

RESPUESTA DE LA IA

«Basta entrevistar 278 voluntarios para garantizar un error del 5 %.»

Tu tarea. Localiza el error, contrástalo con el procedimiento de la unidad (supuestos de la fórmula, 8.1, 8.2 y 8.5) y escribe una respuesta corregida que diga qué permitirían afirmar 278 voluntarios y cuál es su límite. Conserva el cálculo o la definición que justifica cada corrección.

Ojo con la IA. Una cifra correcta con supuestos falsos produce una conclusión falsa: pide siempre las condiciones de la fórmula.

ACTIVIDADES FINALES

Comprueba

12. Resume una distinción decisiva de esta unidad (error de muestreo y sesgo, respuestas y contactos, o población grande y finita) y aplícala al caso resuelto.
13. ¿Por qué con p = q = 0,5 se obtiene el tamaño mayor? Compruébalo con N = 1.000 y p = 0,3.
14. Explica por qué 278 respuestas dan ±5 puntos en el caso y en la oferta de Quique 500 respuestas no garantizan ±4,4.

Aplica

15. Repite la práctica de Ruedalia cambiando una condición relevante (por ejemplo, ±4 % en lugar de ±5 %, una respuesta prevista del 50 % o una base de 2.000 clientes). Señala qué resultado cambia y cuál se mantiene.
16. Nora pide resultados por zona con ±10 puntos en cada una. Calcula, con la fórmula finita, cuántas respuestas necesitaría el Sur (230 hogares) y compara con las 64 de la afijación proporcional.
17. Con la tabla 5, calcula cuántos contactos más harían falta en el Norte si se mantuviera su tasa observada, y qué pasaría con el reparto si, en su lugar, se ponderara.

Investiga

18. Localiza una fuente primaria o documentación técnica sobre no respuesta y limitaciones (la metodología de una encuesta del INE, un manual universitario de muestreo o la guía de un organismo estadístico). Registra la fecha de consulta, el ámbito, la afirmación que sustenta y una limitación.
19. Busca en la ficha técnica de una encuesta publicada el tamaño, el error de muestreo y el nivel de confianza que declara, y comprueba con la fórmula de 8.1 si son coherentes. Anota los supuestos que da (o que faltan).

TU INVESTIGACIÓN AVANZA · PASO 8 · I08 · TAMAÑO Y CONTACTOS

Para la investigación de tu negocio, con el marco de tu I07: 1) supuestos escritos (N, diseño, confianza, error y p, con su justificación); 2) cálculo en hoja de cálculo, con numerador, denominador y redondeo al alza; 3) error comprobado con el tamaño redondeado; 4) reparto por estratos, si los hay, con la suma comprobada; 5) contactos con la tasa de respuesta prevista y su fuente; 6) registro de contactos preparado; 7) limitaciones (cobertura y no respuesta), responsable y fecha de revisión.
Comprueba tu paso: cualquiera puede repetir tu cálculo con tu archivo; ningún contacto de más se presenta como corrección del sesgo.
☐ Hecho ☐ Revisado ☐ Guardado

OJO PROFESIONAL

Documento: borrador del plan de tamaño y contactos EK-2027/TM-v0.1

Documento ficticio, preparado por Unai el martes 20/04/2027 a partir de la oferta de Quique, para que lo revise el equipo de análisis. Contiene **cinco errores**.

N.º	Campo	Contenido
1	Marco	Listado de contactos autorizados 2027, LC-v1.0: N = 1.000 contactos
2	Supuestos	Aleatorio simple sin reposición; 95 % (z = 1,96); p = q = 0,5; e = 0,05
3	Tamaño	n = 277,74 → 277 respuestas completas
4	Contactos	Llamaremos a 278 hogares: 278 contactos garantizan 278 respuestas
5	Si faltan respuestas	Se completan con voluntarios de las redes: cuantos más voluntarios, menos sesgo
6	Más confianza	Si Nora pide el 99 %, se mantiene z = 1,96 y el tamaño no cambia
7	Conclusión prevista	Los resultados valdrán para los 6.000 hogares del barrio con un error de ±5 %
8	Reparto	Norte 117 · Centro 97 · Sur 64

Encargo. Señala los cinco errores, explica por qué lo son con lo que has visto en la unidad y reescribe cada línea afectada para que el plan se pueda usar. **Para pensar:** ¿qué línea correcta solo cuadra si se corrige antes otra?

N.º	Línea	Por qué es un error	Corrección
1			
2			
3			
4			
5			

AUTOEVALUACIÓN

1. Con N = 1.000, e = 5 % y p = q = 0,5, se pasa del 95 % al 99 %. El tamaño...
a) baja a 214.
b) sube a 400.
c) no cambia: N es el mismo.

2. La corrección para población finita cambia mucho el tamaño cuando...
a) la muestra es una parte apreciable de la población.
b) la población es muy grande.
c) p = 0,5.

3. Hacen falta 376 respuestas completas y se prevé un 70 % de respuesta. Hay que contactar a...
a) 376 hogares.
b) 537 hogares.
c) 538 hogares.

4. Afijación proporcional de 278 respuestas en zonas de 420, 350 y 230 hogares. Al Sur le corresponden...
a) 64.
b) 63, la parte entera.
c) 93, como a las demás.

5. 399 contactos dan 278 completas, pero el Norte responde un 60,1 % y el Sur un 79,3 %. Lo correcto es...
a) darlo por bueno: hay 278.
b) comparar perfiles, reintentar en el Norte y ponderar si hace falta.
c) completar el Norte con voluntarios.

6. N = 1.000; 95 %; p = q = 0,5; e = 0,05; respuesta prevista del 70 %. Respuestas y contactos:
a) 277 y 396.
b) 278 y 398.
c) 385 y 550.

7. Ruedalia: N = 500 con los mismos supuestos. Respuestas y contactos:
a) 217 y 310.
b) 385 y 550.
c) 218 y 312.

8. «Basta entrevistar 278 voluntarios para garantizar un error del 5 %» es falso porque...
a) para el 5 % siempre hacen falta 385.
b) la fórmula supone un sorteo; con voluntarios no hay probabilidad conocida.
c) con voluntarios hay que llamar a 398.

9. En el Ojo profesional, «278 contactos garantizan 278 respuestas» se corrige...
a) contactando a 278 y sustituyendo a quien no responda.
b) bajando el tamaño a 195, el 70 % de 278.
c) previendo 398 contactos al 70 % y registrando la respuesta real.

10. En tu I08, ¿qué dato te haría revisar el plan de contactos?
a) Una tasa de respuesta inicial muy inferior a la prevista o desigual entre estratos.
b) La fecha en el nombre del archivo.
c) Un tamaño par.

Ya sé...

- ☐ explicar cómo influyen confianza, error, variabilidad, población y diseño, y calcular el error de un tamaño dado.
- ☐ calcular el tamaño con la fórmula de población finita, redondear al alza y comprobar el error.
- ☐ transferir el procedimiento a otro negocio, repartirlo por estratos y planificar los contactos.
- ☐ detectar una conclusión no sustentada: precisión sin selección o contactos sin control del sesgo.
- ☐ entregar un plan de tamaño y contactos revisable, con sus supuestos y limitaciones.



UNIDAD 9

Depuración, codificación y descripción

Bloque 3 · Muestreo y preparación de datos

EN ESTA UNIDAD APRENDERÁS A…

- Elaborar un diccionario de datos con el tipo, el código, la unidad y el rango de cada variable, y distinguir «no contesta», «no procede» y cero.
- Depurar un registro con reglas escritas: duplicados, filtros, códigos fuera de rango y valores extremos comprobados en el original.
- Construir tablas de frecuencias con su base y elegir el gráfico según la variable y la pregunta.
- Calcular media, mediana, moda, rango y desviación típica muestral con fórmulas visibles en una hoja de cálculo.
- Describir la forma de una distribución con un gráfico y un coeficiente cuya convención se nombra.

SITUACIÓN DE PARTIDA

Lunes 3 de mayo de 2027. La encuesta telefónica de seguimiento a los hogares que probaron los kits en abril ha empezado bien: Adrián ha pasado a la hoja de cálculo las ocho primeras fichas en papel y la variable que más interesa a Itxaso, los **kits del último pedido**, ya tiene una media. Pero Adrián frunce el ceño: hay un hogar que pidió 9 kits.

- Unai propone borrar esa fila: «Es un caso raro; si lo dejamos, el informe sale inflado».
- Itxaso quiere saber cuántos kits lleva un pedido normal para preparar las cajas isotérmicas de los repartos.
- Nora necesita la descripción el viernes 14 de mayo y no quiere decidir con un número que nadie sepa explicar.

Nora te encarga que revises el registro, escribas las reglas de depuración y prepares una descripción comprobable.

PARA EMPEZAR

1. ¿Qué debe decidir o comprobar Nora antes de usar la media de kits por pedido para preparar las cajas?
2. ¿Qué sabemos del pedido de 9 kits y qué estamos suponiendo cuando decimos que «es un caso raro»?
3. ¿Qué prueba permitiría revisar la primera respuesta, la de borrar la fila? ¿Y si el hogar no hubiera contestado: sería lo mismo que cero kits?



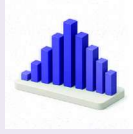
Diccionario

9.1



Limpieza

9.2



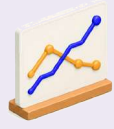
Frecuencias

9.3



Centro y dispersión

9.4



Forma

9.5

9.1 Diccionario y validación

Antes de calcular nada, hay que saber qué significa cada celda.

De la plantilla al diccionario. Ya sabes codificar respuestas en una plantilla (1 = tienda, 2 = web) y apartar los cuestionarios incompletos. Cuando otras personas van a usar, revisar o repetir la investigación, hace falta además un **diccionario de datos**: el documento que describe cada variable con su nombre, su tipo, su unidad, sus códigos o su rango válido y la regla con la que se comprueba. Quien graba, quien depura y quien analiza leen lo mismo en cada celda. Se versiona como cualquier documento del estudio: si cambia un código, cambia la versión y se anota el motivo.

Qué es una fila. La primera línea del diccionario declara la **unidad de análisis**. En la encuesta de seguimiento, una fila es **una entrevista a un hogar**, identificado con un código y nunca con el nombre. Si un hogar aparece dos veces, la regla «una fila por hogar» obliga a revisarlo. En otra base la fila puede ser el pedido, el tique o la respuesta a una pregunta: de esa decisión dependen el recuento y los denominadores.

Tipos de variable. El tipo decide la tabla, el gráfico y las medidas. Las **cualitativas** son **nominales** si sus categorías no tienen orden (canal: tienda, web), y entonces se cuentan pero no se promedian, u **ordinales** si lo tienen (valoración de 1 a 5), que admiten mediana. Las **cuantitativas** son **discretas** si se cuentan (kits, pedidos) y **continuas** si se miden (importe en euros, minutos).

Codificar las respuestas abiertas. «¿Por qué eligió ese kit?» llega en texto libre. Antes de contar se crea una lista de **categorías exhaustivas y excluyentes** (motivo principal: precio, comodidad, horario, producto, otros), cada una con su código y una definición, y se prueba con las primeras respuestas: si «otros» pasa de una de cada diez, se revisa la lista. Dos personas codifican una parte por separado y comparan: las discrepancias afinan las definiciones.

Documento 1 · Diccionario de datos EK-2027/DD-v1.0 (extracto)

(ficticio)

Variable	Tipo y unidad	Códigos o rango	Regla de validación
hogar	texto (clave)	H-001 a H-100	existe en el marco; único
kits_ult	discreta, kits	1 a 20; 98; 99	entero; 98 si P7 = No
canal_ult	nominal	1 tienda; 2 web; 98; 99	98 si P7 = No
pedidos_abr	discreta, pedidos	0 a 30; 99	0 es un valor válido

Ausencia, no procede y cero. Son tres cosas distintas:

- **No contesta (99):** la pregunta se aplicaba y no hay respuesta. Es una **ausencia**: no se sustituye por cero ni por la media; se cuenta y se informa.
- **No procede (98):** el filtro salta la pregunta (quien nunca ha pedido no tiene «último pedido»). Ese hogar queda fuera de la base de la variable.

- **Cero:** el hogar contestó y el valor es 0. En «pedidos en abril» es una respuesta válida que cuenta en la media; en «kits del último pedido», es imposible para quien ha pedido y se revisa.

Los códigos 98 y 99 son **etiquetas**, no cantidades. Promediados no significan nada: con las ocho filas del registro, (2 + 4 + 4 + 4 + 6 + 99 + 9 + 98) ÷ 8 = 226 ÷ 8 = **28,25 kits** por pedido. Con R04 ya fuera (v2), una condición: =PROMEDIO.SI(C2:C8;"<98") → 5; en el bruto (C2:C9), 4,83.



Validación. Lo más barato es que el error no entre: la validación de datos de la columna kits_ult admite solo enteros de 1 a 20, 98 y 99, y en el formulario la pregunta solo aparece si P7 = Sí. Después de grabar se comprueban **tipo**, **rango**, **coherencia** (si P7 = No, kits_ult y canal_ult valen 98) y **unicidad**. Los datos en bruto se guardan aparte y sin tocar (`2027-05-03_seguimiento_kits_brutos_v1`); la depuración trabaja sobre una copia versionada.

LA NORMA DICE

Reglamento (UE) 2016/679 (RGPD), de 27 de abril de 2016, art. 5.1: los datos personales serán «adecuados, pertinentes y limitados a lo necesario en relación con los fines para los que son tratados» (c, minimización) y «exactos y, si fuera necesario, actualizados» (d, exactitud). Por eso la base usa el código H-042 y no el nombre, y un dato inexacto se rectifica con su fuente. Consulta la versión vigente.

ACTIVIDADES

1. Crea el diccionario, limpia la tabla y produce las frecuencias con un gráfico adecuado. Registro ficticio de valoraciones (V01 a V06): kit K02, K03, K02, K04, K02, K03; valoración (1 nada, 5 mucho; 9 = no sabe) 4, 5, 9, 3, 4, 7; pedidos en abril 2, 0, 1, 1, 3, 2. Di qué haces con el 9, el 7 (sin ficha) y el 0, y la base de cada tabla.
2. Clasifica como no contesta, no procede, cero válido o error: a) un hogar que nunca ha pedido, en kits_ult; b) un hogar que no pidió en abril, en pedidos_abr; c) un hogar que colgó antes de la pregunta de kits; d) kits_ult = 0 con P7 = Sí.
3. Escribe las reglas de validación de canal_ult (lista admitida y coherencia con P7) y explica qué error evita cada una.

9.2 Ausencias, duplicados y filtros

Depurar no es dejar los datos como nos gustan: es dejarlos como son, y explicar cada cambio.

El registro bruto. Estas son las ocho primeras fichas, tal como Adrián las grabó el lunes 3 de mayo; la columna de observaciones copia lo que anotó el entrevistador en el papel:

Registro 1 · Encuesta de seguimiento, kits del último pedido (en bruto)

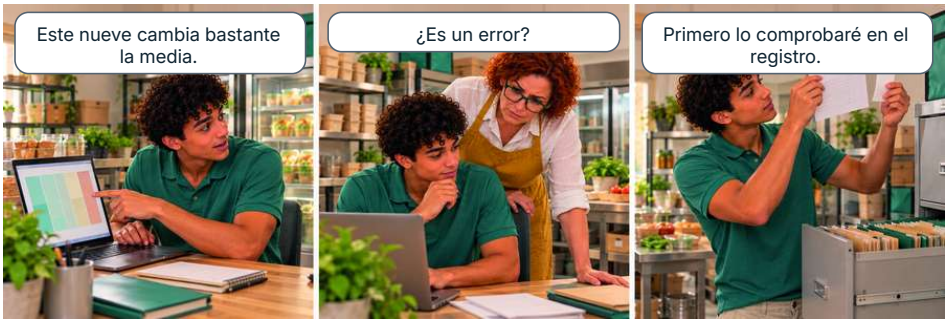
(ficticio; 98 = no procede; 99 = no contesta)

Registro	Hogar	kits_ult (kits)	Observación de la ficha
R01	H-003	2	
R02	H-011	4	
R03	H-017	4	03/05, 10:12
R04	H-017	4	03/05, 10:12
R05	H-024	6	
R06	H-030	99	colgó antes de la pregunta
R07	H-042	9	
R08	H-051	98	P7 = No (nunca ha pedido)

Reglas de depuración. Se escriben antes de mirar los resultados y se aplican en orden sobre la copia:

1. **Duplicados:** mismo hogar y hora = la misma ficha grabada dos veces; se conserva la primera. Una columna de control, =CONTAR.SI.CONJUNTO($B$2:B2;B2;$D$2:D2;D2)>1, marca la segunda si D guarda fecha y hora (aquí, la observación). Otra fecha: se consulta.
2. **No procede:** las filas con 98 salen de la base de la variable y se cuentan aparte.
3. **No contesta:** las filas con 99 no se rellenan; se informa de la **tasa de respuesta de la variable**.
4. **Coherencia y rango:** lo que no cumple el diccionario se contrasta con la ficha en papel.
5. **Extremos:** un valor dentro del rango pero alejado del resto **se comprueba en el original**. Si es un error de grabación, se corrige con la fuente; si es real, se conserva; si no se puede comprobar, se conserva y se analiza con él y sin él.

Un extremo no es un error. El hogar H-042 declaró 9 kits. Una regla habitual usa los **cuartiles**: con =CUARTIL.INC, Q1 = 4 y Q3 = 6; el **rango intercuartílico** es RIC = 2, y se revisan los valores fuera de Q1 − 1,5·RIC = 1 y Q3 + 1,5·RIC = 9. El 9 queda justo en el límite, no fuera: la regla solo avisa. Adrián lo revisa porque cambia mucho la media, y eso basta. Itxaso localiza el tique **T-2027/04415**: pedido real de 9 kits para una comida familiar. Se conserva.



Registro 2 · Registro de depuración EK-2027, v1 → v2

(ficticio)

N.º	Fila	Regla aplicada	Acción y evidencia
1	R04	Duplicado de R03 (hogar y hora)	Fuera de la v2; sigue en el bruto
2	R08	No procede (P7 = No)	Fuera de la base de kits_ult
3	R06	No contesta	Ausencia; se informa
4	R07	Extremo comprobado	Se conserva: tique T-2027/04415

Resultado. Quedan **2, 4, 4, 6, 9** (n = 5). Los denominadores se dicen siempre: 8 registros, 7 hogares distintos, 6 hogares a los que se aplicaba la pregunta y 5 respuestas válidas. Tasa de respuesta de la variable: 5 ÷ 6 = **83,3 %**. Si el hogar que colgó fuera de los que piden mucho, la descripción cambiaría: por eso la ausencia se informa.

¿Y rellenar la ausencia con la media? Inventa un dato. Con el 99 de R06 sustituido por 5, los datos serían 2, 4, 4, 5, 6, 9: la media seguiría en 5, pero s bajaría de 2,646 a √(28 ÷ 5) ≈ 2,366 kits. La hoja parecería más precisa sin tener más información.

CLAVE · NO BORRAR UN DATO VÁLIDO PORQUE ESTORBA

Un dato no se elimina porque cambie el resultado esperado, sino porque una regla escrita y una comprobación demuestran que es un error. Cada cambio queda registrado con fila, regla, acción y evidencia.

ACTIVIDADES

4. Unai insiste en borrar R07 «porque es atípico». Redacta la respuesta del equipo: regla aplicada, evidencia, una alternativa razonable (describir con el 9 y sin él) y la condición que te haría cambiar de decisión. Añádelo a tu ficha de la entrega I09.
5. Supón que R04 tuviera la hora 18:40 del miércoles 5 de mayo. ¿Sería un duplicado? ¿Qué harías? Comprueba además que la tasa de respuesta es del 83,3 % y explica por qué no se calcula sobre los 8 registros.

9.3 Frecuencias y gráficos

Un porcentaje sin su base es una opinión con decimales.

Tabla de frecuencias.

La **frecuencia absoluta** (n_i) es el número de casos de cada valor; la **relativa** ($f_i = n_i \div n$), su proporción sobre la base, que por 100 es un porcentaje; la **acumulada** suma las anteriores y solo tiene sentido si los valores tienen orden.

Tabla 1 · Kits del último pedido

(base: 5 hogares con dato; encuesta de seguimiento, mayo de 2027; ficticia)

Kits	Hogares (n_i)	% (f_i)	% acumulado
2	1	20,0	20,0
4	2	40,0	60,0
6	1	20,0	80,0
9	1	20,0	100,0
Total	**5**	**100,0**	

El 60,0 % de los hogares con dato pidió 4 kits o menos. Con cinco casos, cada hogar pesa un 20 %: junto al porcentaje se dan las frecuencias absolutas («3 de 5»), que no esconden lo pequeña que es la base.

Bases compatibles. Cada porcentaje se calcula sobre la base de su pregunta. El canal del último pedido va sobre los 5 hogares con dato: web 3 (60,0 %) y tienda 2 (40,0 %). Quien escribe «web: 37,5 %» ha dividido 3 entre los 8 registros, que incluyen un duplicado, una ausencia y un hogar sin pedidos: no es el porcentaje de nada. La calidad del trabajo de campo va en otra tabla con otra base: de 7 hogares, 5 con dato (71,4 %), 1 no contesta (14,3 %) y 1 no procede (14,3 %).

Tabla 2 · Qué gráfico para qué variable

Variable	Pregunta	Gráfico adecuado	Evita
Nominal u ordinal	¿qué categoría pesa más?	barras (orden de la escala si es ordinal)	sectores con muchas porciones
Discreta, n pequeño	¿dónde se concentran?	diagrama de puntos	histograma de 5 casos
Continua	¿qué forma tiene?	histograma con intervalos iguales	barras sueltas por valor
Dos cuantitativas	¿van juntas?	dispersión	dos gráficos separados

Un gráfico que se pueda leer. Título informativo, unidad en el eje, **base** (n) y fuente con fecha; en barras, eje desde cero. Los sectores solo valen para partes de un todo con pocas categorías: con dos, basta una frase («3 de 5 pidieron por la web»). La tabla dinámica o =CONTAR.SI(C2:C6;4) da las frecuencias, y el gráfico se crea desde esa tabla para que se actualice.

EN TU SECTOR

Marketing y Publicidad: el recuerdo de un anuncio se presenta en barras con «base: personas que lo vieron», no sobre toda la muestra. **Gestión de Ventas y Espacios Comerciales:** los artículos por tique se presentan en barras por valor, con la base de tiques del periodo, para dimensionar cajas o puestos de cobro.

ACTIVIDADES

6. Elabora la tabla de frecuencias del canal del último pedido de los cinco hogares y elige un gráfico. Entrega el ejemplo documentado (tabla, base y fuente), una alternativa de presentación y la condición que te haría cambiar de gráfico. Usa la ficha de la entrega I09.
7. Un borrador dice: «Canal: web 37,5 %; tienda 25,0 %; resto, sin dato». Explica qué base se ha usado, por qué es incorrecta y reescribe la línea.

9.4 Tendencia central y dispersión

Una media sola no describe nada: hace falta el centro y cuánto se separan los datos de él.

Tendencia central. Con los datos depurados (2, 4, 4, 6, 9 kits; n = 5):

- **Media:** $\bar{x} = \Sigma x \div n = (2 + 4 + 4 + 6 + 9) \div 5 = 25 \div 5 =$ **5 kits**. Usa todos los valores y es sensible a los extremos.
- **Mediana:** valor central de los datos ordenados; con n impar, posición $(n + 1) \div 2 = 3$: **4 kits** (con n par, media de los dos centrales). Resiste mejor los extremos, pero no es inmune: sin el 2 (4, 4, 6, 9) pasaría a $(4 + 6) \div 2 = 5$.
- **Moda:** valor más frecuente: **4 kits**. Es la única medida de centro para una variable nominal.

Dispersión. El **rango** es máximo menos mínimo: 9 − 2 = **7 kits**; solo mira dos datos. La **desviación típica** resume cuánto se separan los valores de la media, en la unidad de los datos.

Ejemplo resuelto. Desviación típica muestral de los cinco pedidos

Hogar	x (kits)	$x - \bar{x}$	$(x - \bar{x})^2$
H-003	2	−3	9
H-011	4	−1	1
H-017	4	−1	1
H-024	6	1	1
H-042	9	4	16
Suma	**25**	**0**	**28**

Comprobación: las desviaciones suman 0. Varianza muestral: $s^2 = \Sigma(x - \bar{x})^2 \div (n - 1) = 28 \div 4 = 7$; desviación típica muestral: $s = \sqrt{7} \approx$ **2,646 kits**. El **coeficiente de variación**, $CV = s \div \bar{x} = 2{,}646 \div 5 =$ **52,9 %**, relaciona la dispersión con la media y permite comparar variables de distinta unidad.

s con n – 1 y σ con N. Si los datos son una **muestra** y describes la variabilidad de la población de la que salen, divides entre n – 1 (**desviación típica muestral**, s): las desviaciones se miden respecto a la media de la propia muestra, y dividir entre n subestimaría la variabilidad. Si son **toda la población** de interés (por ejemplo, los cinco pedidos de una semana y nada más), divides entre N (**desviación típica poblacional**, σ). Los cinco hogares son una parte de los que probaron los kits: s ≈ 2,646. Con N saldría σ = √(28 ÷ 5) = √5,6 ≈ 2,366, que aquí infravalora la dispersión. Escríbelo siempre: «s (muestral, n – 1)» o «σ (poblacional, N)». En la hoja, ambas fórmulas existen y dan resultados distintos con los mismos datos.

Diagrama de puntos y comparación entre media y mediana (recurso editable: un punto por hogar)

Kits del último pedido	Hogares	Medidas
2	●	
3		
4	● ●	mediana y moda (4)
5		media (5)
6	●	
7-8		
9	●	máximo, comprobado

Tres de los cinco pedidos están en 4 o menos y el de 9 «tira» de la media. Sin el 9, la media bajaría a 4 y coincidiría con la mediana: el extremo mueve la media un kit entero y no mueve la mediana.

En la hoja de cálculo, con fórmulas visibles. Datos depurados en B2:B6: =PROMEDIO(B2:B6) → 5; =MEDIANA(B2:B6) → 4; =MODA.UNO(B2:B6) → 4; =MAX(B2:B6)-MIN(B2:B6) → 7; =DESVEST.M(B2:B6) → 2,646 (n – 1); =DESVEST.P(B2:B6) → 2,366 (N). Los nombres cambian algo entre programas: comprueba en la ayuda cuál divide entre n – 1, y añade una celda de control que recalcule s a mano.

Comprobar que la hoja se actualiza. Si el tique de H-042 hubiera dicho 3 kits (error de grabación), los datos serían 2, 3, 4, 4, 6: media 3,8; mediana 4; rango 4; s = √2,2 ≈ 1,483 kits. Si algún resultado no cambia, hay un número tecleado donde debería haber una fórmula.

ACTIVIDADES

8. Calcula media, mediana, moda, rango y desviación típica muestral de seis pedidos de oficinas de la prueba (kits por pedido, ficticios): 3, 5, 5, 6, 8, 12. Calcula también σ y di cuándo usarías cada una. Analiza la asimetría con el diagrama de puntos y con el coeficiente de la hoja, y explica qué diría la curtosis y por qué con seis datos no conviene interpretarla.

9. Entrega la hoja de los cinco hogares con datos, fórmulas visibles, celda de control y gráfico. Después cambia el 9 por un 3, comprueba que se actualizan todos los resultados y anota los valores antes y después.

9.5 Forma de la distribución

Media 5 y mediana 4 sugieren una cola a la derecha. Sugerir no es demostrar.

Simetría, colas y curtosis. Una distribución es **simétrica** si los valores se reparten igual a los dos lados del centro; tiene **asimetría a la derecha** (positiva) si se alarga una cola de valores altos, y **a la izquierda** (negativa) si la cola es de valores bajos. En los pedidos es habitual la cola derecha: muchos pedidos medianos y unos pocos grandes. La **curtosis** describe el peso de las colas frente a la distribución normal: una curtosis alta indica más valores muy alejados de lo esperable.

Tres señales que se miran juntas: el gráfico, la posición de la media respecto a la mediana y un coeficiente cuya convención se nombra. Ninguna basta sola: hay distribuciones asimétricas con media y mediana iguales, y un coeficiente alto con cinco datos puede deberse a un único valor.

Coeficiente de asimetría con su convención. Hay varias fórmulas y dan valores distintos:

- **g1 (de momentos):** $g1 = m_3 \div m_2^{(3/2)}$, con $m_2 = \Sigma(x - \bar{x})^2 \div n$ y $m_3 = \Sigma(x - \bar{x})^3 \div n$. En el caso: $m_2 = 28 \div 5 = 5{,}6$; $\Sigma(x - \bar{x})^3 = -27 - 1 - 1 + 1 + 64 = 36$, $m_3 = 7{,}2$; $g1 = 7{,}2 \div 13{,}252 \approx$ **0,543**.
- **G1 (ajustado por tamaño, el de la hoja de cálculo,** =COEFICIENTE.ASIMETRIA): $G1 = g1 \cdot \sqrt{[n(n-1)]} \div (n-2) = 0{,}543 \cdot \sqrt{20} \div 3 \approx$ **0,810**.

Los dos son positivos: cola a la derecha. El informe no dice «asimetría 0,8», sino «G1 (hoja de cálculo) ≈ 0,810, n = 5». La **curtosis** de la hoja (=CURTOSIS, exceso sobre la normal, convención G2) da ≈ 0,673; con cinco datos es muy inestable y no se interpreta.

Sensibilidad. Sin el 9 (2, 4, 4, 6), la media y la mediana valen 4 y G1 = 0: toda la asimetría la produce un solo pedido, comprobado y real. La descripción honrada es: «cola derecha por un pedido grande real; el pedido típico es de 4 kits; con cinco hogares, la forma es orientativa». Con cola derecha, la mediana describe el pedido **típico** y la media sirve para **totales** (cinco pedidos de media 5 son 25 kits que preparar). Itxaso necesita las dos, y además el máximo comprobado.

PARA SABER MÁS

El **diagrama de caja** resume la forma con mínimo, Q1, mediana, Q3 y máximo. Di qué convención de cuartiles usas.

ACTIVIDADES

10. Redacta la descripción de la forma de los cinco pedidos para el informe de Nora: gráfico, media y mediana, coeficiente con su convención y límite por el tamaño. Entrega un ejemplo documentado, una alternativa (describir sin el 9) y la condición que haría cambiar tu descripción. Usa la ficha de la entrega I09.
11. Calcula g1 y G1 de los pedidos de oficinas de la actividad 8 y explica por qué no puedes concluir solo con «la media es mayor que la mediana».

CASO RESUELTO · PASO A PASO

Las cinco compras de la prueba

Caso ficticio. Viernes 14 de mayo de 2027. El equipo de análisis entrega a Nora la descripción de los kits del último pedido, con los datos depurados y su registro de cambios.

Paso 1. Decisión, datos y alcance. Decisión: qué tamaño de caja isotérmica estándar prepara Itxaso para el reparto a hogares y cuándo hace falta una segunda caja. Datos: encuesta de seguimiento, kits del último pedido (unidades), 8 registros → 5 hogares con dato válido tras la depuración (registro 2). Alcance: **estos cinco hogares**, una muestra de los que probaron los kits; se describen sin estimar con precisión la media de todos ni mezclarlos con otros estudios.

Paso 2. Procedimiento, con unidades y condiciones. Datos ordenados: 2, 4, 4, 6, 9 kits.

Medida	Cálculo	Resultado
Media	25 ÷ 5	5 kits
Mediana	posición (5 + 1) ÷ 2 = 3.ª	4 kits
Moda	valor que se repite (dos veces)	4 kits
Rango	9 − 2	7 kits
s (muestral, n − 1)	√(28 ÷ 4) = √7	≈ 2,646 kits
Forma	G1 de la hoja, n = 5	≈ 0,810 (cola derecha)

Condiciones: códigos 98 y 99 fuera del cálculo; duplicado R04 excluido; extremo R07 comprobado en el tique T-2027/04415 y conservado.

Paso 3. Comprobación e interpretación. El diagrama de puntos (9.4) muestra tres pedidos en 4 o menos y uno aislado en 9. Media mayor que mediana y G1 positivo coinciden: cola derecha. La comprobación de sensibilidad (sin el 9: media 4, mediana 4, G1 = 0) indica que toda la asimetría depende de un pedido real. **Qué permite concluir:** en estos cinco hogares el pedido típico es de 4 kits y existen pedidos grandes reales. **Qué no permite:** afirmar que «los hogares piden 5 kits de media», ni estimar un porcentaje de pedidos grandes con cinco casos.

RESULTADO

Media 5 kits; mediana 4; moda 4; rango 7; s = √7 ≈ 2,646 kits (muestral). Distribución con cola derecha: la descripción visual (diagrama de puntos) y el coeficiente formal (G1 ≈ 0,810, hoja de cálculo) coinciden, pero con n = 5 la forma es orientativa. Propuesta para Itxaso: caja estándar para 4 kits y una segunda caja para pedidos mayores, revisable con los datos de la encuesta completa.

Comprueba que lo has entendido. a) ¿Por qué la media de 3,625 kits que saldría con ceros y el duplicado es incorrecta? b) ¿Qué cambiaría si el tique de H-042 dijera 3 kits? c) ¿Por qué no se usa σ = 2,366?

PRÁCTICA

Las revisiones de los clientes de Ruedalia

Datos ficticios. Txema quiere decidir cuántas revisiones incluye la cuota de mantenimiento para particulares en 2028. Ousmane ha sacado de las fichas del taller las revisiones del año 2027 de seis clientes de la cuota: 1, 2, 2, 3, 7 y una ficha sin el dato (no consta, la ficha se perdió en el cambio de programa). Los valores son unidades (revisiones por cliente y año).

Encargo. 1) Escribe el diccionario de la variable (tipo, unidad, rango y código de la ausencia). 2) Depura: decide qué haces con la ficha sin dato y con el 7, y regístralo. 3) Calcula media, mediana, moda, rango y desviación típica muestral en una hoja de cálculo con fórmulas visibles y una celda de control. 4) Haz el diagrama de puntos y describe la forma con su convención. 5) Propón a Txema una decisión con sus limitaciones.

Plantilla IC-P09 · Descripción de las revisiones

Campo	Respuesta del alumno
Datos y fuentes con unidad y periodo	
Procedimiento o cálculo	
Resultado y evidencia	
Decisión y limitaciones	

Lista de cotejo

☐ Ausencia codificada, no convertida en cero.
☐ Base dicha (cuántas fichas y cuántos datos válidos).
☐ s con n − 1, dicho.
☐ Extremo comprobado o analizado con y sin él.
☐ Convención del coeficiente de forma nombrada.
☐ Decisión proporcional a cinco datos.

LECTURA Y DEBATE

Trescientos ochenta euros



Relato ficticio. Miércoles 12 de mayo, a media tarde. Adrián está grabando la segunda tanda de fichas de la encuesta y ya no se fía de su primer impulso. La semana anterior había estado a punto de borrar el pedido de 9 kits, y resultó real. Por eso, cuando en la columna del importe del último pedido aparece **380 €** al lado de 21, 32, 38 y 44, piensa: «Será otra comida familiar», y lo deja.

Itxaso pasa por detrás, con una caja isotérmica, y mira la pantalla. «¿Trescientos ochenta? Eso son unos 30 kits. No ha salido un pedido así en todo el mes». Adrián le explica que la última vez el dato grande era verdad. «Precisamente», contesta ella, «la última vez lo comprobaste. Hoy lo estás suponiendo».

Abren juntos el archivo de albaranes de la zona de kits. El pedido del hogar es de 38,00 €: dos kits K02 y un K05, más que razonable. En la ficha en papel, el 3 y el 8 están escritos con prisa y un trazo del bolígrafo parece un cero. Adrián corrige la celda, anota en el registro de depuración «importe: 380 → 38,00, error de grabación, albarán de la zona de kits» y recalcula: la media del importe baja de 103,00 € a 34,60 €, y la mediana sigue en 38 €.

«Ni borrar ni conservar por costumbre», resume Itxaso. «Comprobar».

Preguntas. 1) ¿Qué afirmación de la historia necesita una comprobación antes de convertirse en decisión? 2) ¿Qué información pedirías primero y por qué cambiaría la respuesta? 3) Propón una actuación provisional para los importes que no se puedan comprobar hoy, que permita seguir trabajando sin ocultar la incertidumbre.

APRENDE CON LA IA

Unai tiene prisa y pregunta a un asistente de IA si vale la pena calcular la mediana además de la media para el informe de los kits. Esta es la respuesta:

> **RESPUESTA DE LA IA**
> «No hace falta: la mediana siempre coincide con la media y no cambia con ningún dato. Calcula solo la media, que es la medida estándar.»

Tu tarea. Localiza el error (en realidad hay dos afirmaciones falsas en una frase), contrasta cada una con los datos de la unidad (2, 4, 4, 6, 9 y la sensibilidad de 9.5) y escribe una respuesta corregida con su límite: qué mide cada medida, cuándo conviene cada una y qué no garantiza la mediana. Conserva el cálculo que justifica la corrección.

Ojo con la IA. Las respuestas rotundas («siempre», «nunca», «no cambia») son las primeras que hay que comprobar con un contraejemplo. No pegues en el asistente datos de hogares identificables.

ACTIVIDADES FINALES

Comprueba

12. Resume una distinción que sea decisiva en esta unidad (ausencia, no procede y cero; o s y σ) y aplícala al caso resuelto: qué resultado saldría si no la respetaras.
13. Explica qué representa una fila en la encuesta de seguimiento y por qué R04 no es «otra respuesta».
14. ¿Por qué la tasa de respuesta de kits_ult se calcula sobre 6 hogares y no sobre 8 registros ni sobre 7 hogares?
15. Ordena de más a menos sensible a un valor extremo: rango, mediana, media. Justifícalo con los datos del caso.

Aplica

16. Repite la práctica de Ruedalia cambiando una condición relevante: Ousmane encuentra que el 7 era un 1 mal grabado. Señala qué resultados cambian (media, moda, desviación, forma) y cuál se mantiene (mediana), y si cambia la decisión de Txema.
17. Repite la práctica suponiendo ahora que se recupera la ficha perdida y el cliente hizo 2 revisiones. ¿Qué medidas cambian? ¿Cambia la base?
18. Los importes corregidos de la lectura son 21, 32, 38, 38 y 44 € sin IVA. Calcula media, mediana y desviación típica muestral, y di si la variable es discreta o continua.

Investiga

19. Localiza una fuente primaria o documentación técnica sobre la forma de la distribución (la ayuda oficial de tu hoja de cálculo para COEFICIENTE.ASIMETRIA y CURTOSIS, o un manual universitario). Registra fecha, ámbito, la afirmación que sustenta (qué fórmula usa) y una limitación.
20. Busca en la ayuda de tu hoja de cálculo cómo se define CUARTIL.INC y en qué se diferencia de CUARTIL.EXC. Calcula los dos con 2, 4, 4, 6, 9 y anota la diferencia.

TU INVESTIGACIÓN AVANZA · PASO 9 · I09 · DATOS DEPURADOS Y DESCRIPCIÓN

Con los datos ficticios o propios autorizados de tu investigación: 1) diccionario de datos con unidad de análisis, tipo, unidad, códigos y reglas de validación; 2) datos en bruto intactos y copia depurada con versión; 3) registro de depuración (fila, regla, acción y evidencia); 4) tabla de frecuencias con base y un gráfico adecuado; 5) media, mediana, moda, rango y s con fórmulas visibles y celda de control; 6) forma con gráfico y coeficiente con su convención; 7) decisión, responsable y fecha de revisión. Modelo: el caso de los cinco pedidos (2, 4, 4, 6, 9; s = √7).

Comprueba tu paso: ningún código convertido en cero, denominadores visibles y la hoja se recalcula al cambiar un dato. No basta una presentación sin el archivo de cálculo.

☐ Hecho ☐ Revisado ☐ Guardado

OJO PROFESIONAL

Documento: nota de resultados de la encuesta de seguimiento

Documento ficticio, preparado por Unai el martes 4 de mayo de 2027 para el informe de La Despensa de Nora, S.L., antes de la depuración del equipo. Contiene **cinco errores**.

N.º	Apartado	Contenido
1	Variable	Kits del último pedido (unidades). Encuesta telefónica de seguimiento; 8 registros grabados el 03/05/2027
2	Códigos	«No contesta» (R06) y «no procede» (R08) se han grabado como 0 kits para que la hoja no dé error
3	Duplicados	R04 se conserva: «es otra fila de la hoja»
4	Extremos	R07 (9 kits) se elimina por atípico; no se consulta el tique
5	Medidas	Datos usados: 2, 4, 4, 4, 6, 0, 0. Media 2,86 kits; mediana 4; desviación típica 2,10 kits (=DESVEST.P), «la variabilidad de los hogares que prueban los kits»
6	Canal	Web: 3 de 8 registros (37,5 %)
7	Archivo	Datos en bruto guardados sin tocar en `2027-05-03_seguimiento_kits_brutos_v1`
8	Conclusión	«El hogar típico pide menos de 3 kits: basta una caja pequeña»

Encargo. Anota las cinco correcciones justificadas (línea, qué falla y cómo debe quedar) y reescribe la nota para que Nora la pueda usar: datos depurados, medidas con su divisor, base de cada porcentaje y una conclusión con su límite. **Para pensar:** dos líneas son correctas y una tercera no tiene un error propio: es la consecuencia de los anteriores. ¿Cuáles son?

N.º	Línea	Por qué es un error	Corrección
1			
2			
3			
4			
5			

AUTOEVALUACIÓN

1. En kits_ult, un hogar que nunca ha pedido se codifica...
a) 0, porque no pidió ningún kit.
b) 98 (no procede), y queda fuera de la base de la variable.
c) 99 (no contesta), y se cuenta como ausencia.

2. Unai quiere borrar un pedido de 9 kits que está en el tique T-2027/04415. Lo correcto es...
a) borrarlo, porque sube la media.
b) sustituirlo por la mediana.
c) conservarlo, porque está comprobado, y anotarlo en el registro de depuración.

3. Con 3 pedidos por la web en 5 hogares con dato y 8 registros grabados, el porcentaje de la web es...
a) 60,0 % (base: 5 hogares con dato).
b) 37,5 % (base: 8 registros).
c) 42,9 % (base: 7 hogares).

4. Para 2, 4, 4, 6, 9 (muestra), la desviación típica que procede es...
a) $\sqrt{5,6} \approx 2,366$, con N.
b) $\sqrt{7} \approx 2,646$, con $n - 1$.
c) 7, el rango.

5. Con 2, 4, 4, 6, 9, G1 ≈ 0,810 y media mayor que mediana, lo correcto es decir...
a) «cola derecha, G1 de la hoja con n = 5; sin el 9 sería simétrica».
b) «asimetría 0,8: los hogares piden mucho».
c) «la media mayor que la mediana demuestra la asimetría».

6. La media, la mediana y la moda de 2, 4, 4, 6, 9 son...
a) 4, 4 y 4.
b) 5, 4 y 4.
c) 5, 6 y 4.

7. En Ruedalia (1, 2, 2, 3, 7 y una ficha sin dato), si la ausencia se graba como 0...
a) la media baja de 3 a 2,5 y el resultado es incorrecto.
b) no cambia nada, porque es un solo dato.
c) es correcto, porque el cliente no hizo revisiones.

8. La afirmación «la mediana siempre coincide con la media y no cambia con ningún dato»...
a) es correcta para datos depurados.
b) es falsa: en el caso, media 5 y mediana 4; y sin el 2 la mediana pasa de 4 a 5.
c) es falsa solo con muestras grandes.

9. En la nota de Unai, la media de 2,86 kits es errónea porque...
a) se ha redondeado mal.
b) incluye dos códigos como ceros, un duplicado y excluye un dato real.
c) debería calcularse con DESVEST.P.

10. En tu entrega I09, un dato adicional que te haría revisar la descripción sería...
a) el resto de la encuesta (más hogares) o la ficha del hogar que no contestó.
b) un gráfico con más colores.
c) la media de otro estudio del encargo.

Ya sé...

- ☐ explicar las distinciones de los cinco apartados: diccionario, depuración, frecuencias, centro y dispersión, forma.
- ☐ ejecutar el procedimiento del caso: depurar y describir 2, 4, 4, 6, 9 con $s = \sqrt{7}$.
- ☐ transferirlo a otro negocio, como las revisiones de Ruedalia.
- ☐ detectar una conclusión no sustentada, como una ausencia convertida en cero.
- ☐ entregar una evidencia revisable: hoja con fórmulas, registro de depuración y versión.



UNIDAD 10

Inferencia e interpretación

Bloque 4 · Análisis y comunicación de resultados

EN ESTA UNIDAD APRENDERÁS A…

- Distinguir parámetro, estimador y estimación, y usar la probabilidad y la distribución normal para explicar la variación entre muestras.
- Calcular e interpretar un intervalo de confianza para una proporción, comprobando sus condiciones.
- Separar el error de muestreo de los sesgos de cobertura, selección, medición y no respuesta.
- Comparar grupos, medir la relación entre dos variables y contrastar una hipótesis sin convertir una asociación en una causa.
- Redactar conclusiones con magnitud, incertidumbre y límites que permitan decidir.

SITUACIÓN DE PARTIDA

Lunes 17 de mayo de 2027. Llega el resultado de la encuesta del estudio de kits: de 200 encuestados elegidos por muestreo aleatorio simple, **120** dicen que encargarían el kit con entrega programada en el próximo mes. Unai ya tiene el titular: «¡El 60 % quiere nuestros kits!». Nora lo lee dos veces y hace tres preguntas antes de alegrarse:

- ¿Cómo se eligió a esas 200 personas?
- ¿Cuánto podría cambiar ese 60 % con otra muestra?
- ¿Sirve para decidir el lanzamiento, o solo dice lo que la gente cree que hará?

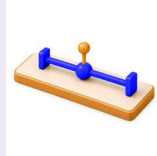
Te pide que el viernes 28 de mayo le entregues una conclusión que se pueda defender.

PARA EMPEZAR

1. ¿Qué debe decidir o comprobar Nora antes de usar el 60 % para lanzar el kit?
2. ¿Qué sabemos (120 de 200, muestra aleatoria) y qué estamos suponiendo (que la intención se convertirá en compra, que quien no contestó piensa igual)?
3. ¿Qué prueba permitiría revisar la primera respuesta: otra muestra, un intervalo, una prueba piloto con ventas reales?



Estimador 10.1



Intervalo 10.2



Error y sesgo 10.3



Comparación 10.4



Conclusión 10.5

10.1 Población, muestra y estimador

El 60 % es de la muestra. Lo que Nora quiere saber está en la población.

Parámetro, estimador y estimación. La **población** es el conjunto completo sobre el que se quiere concluir; aquí, los varios miles de hogares y oficinas del listado del estudio. El **parámetro** es la cifra verdadera de la población, desconocida: p, la proporción que encargaría el kit. El **estimador** es la regla de cálculo que se aplica a la muestra, $\hat{p} = x \div n$ (favorables entre encuestados), y la **estimación** es el valor que da con esta muestra: $\hat{p} = 120 \div 200 =$ **0,60**. El parámetro es fijo; el estimador cambia de una muestra a otra.

Probabilidad: el lenguaje de la incertidumbre. La **probabilidad** de un suceso es un número entre 0 y 1 que mide lo esperable que es: la proporción de veces que ocurriría si repitiéramos el experimento muchas veces. Si eliges al azar una de las 200 respuestas, la probabilidad de que sea favorable es $120 \div 200 = 0{,}6$, y la del suceso complementario, $1 - 0{,}6 = 0{,}4$. Si eliges dos respuestas al azar con reposición, los sucesos son **independientes** y la probabilidad de que las dos sean favorables es $0{,}6 \cdot 0{,}6 = 0{,}36$.

Variabilidad muestral. Con otra muestra de 200 saldría otro $\hat{p}$: 0,57, 0,62, 0,59… Esos valores se reparten alrededor del parámetro con una dispersión que se llama **error estándar** (EE) del estimador. Para una proporción, $EE = \sqrt{p \cdot q \div n}$, con $q = 1 - p$. Cuanto mayor es n, menor es el EE. Con $p = 0{,}6$ y $n = 200$, $EE = \sqrt{0{,}6 \cdot 0{,}4 \div 200} \approx 0{,}035$, unos 3,5 puntos: lo que el azar puede mover $\hat{p}$.

La distribución normal. Muchos fenómenos, y en particular los estimadores calculados con muestras grandes, se reparten según una curva en forma de campana: la **distribución normal**, simétrica alrededor de su media μ y con una dispersión σ. En cualquier normal, entre $\mu - 1\sigma$ y $\mu + 1\sigma$ queda el 68,3 % de los valores; entre $\mu \pm 1{,}96\sigma$, el 95 %; y entre $\mu \pm 2{,}576\sigma$, el 99 %. Para usar una sola tabla se **tipifica**: $z = (\text{valor} - \mu) \div \sigma$ dice a cuántas desviaciones típicas está un valor de la media. En la hoja de cálculo, =DISTR.NORM.ESTAND.N(z;VERDADERO) da la probabilidad acumulada hasta z, y =DISTR.NORM.ESTAND.INV(0,975) devuelve 1,96.

Ejemplo resuelto. ¿Cuánto se mueve $\hat{p}$ si p fuera 0,55? Si en la población el 55 % fuera favorable, con $n = 200$: $EE = \sqrt{0{,}55 \cdot 0{,}45 \div 200} \approx 0{,}0352$. Las condiciones se cumplen ($200 \cdot 0{,}55 = 110 \geq 10$ y $200 \cdot 0{,}45 = 90 \geq 10$), así que $\hat{p}$ sigue aproximadamente una normal y el 95 % de las muestras posibles daría $\hat{p}$ entre $0{,}55 - 1{,}96 \cdot 0{,}0352$ y $0{,}55 + 1{,}96 \cdot 0{,}0352$, es decir, entre el **48,1 %** y el **61,9 %**. Un $\hat{p}$ de 0,62 estaría a $z = (0{,}62 - 0{,}55) \div 0{,}0352 \approx 1{,}99$ errores estándar: posible, pero poco frecuente.

Condiciones de la inferencia. Extrapolar de la muestra a la población exige: 1) **selección aleatoria** (cada unidad con probabilidad conocida; aquí, aleatoria simple); 2) respuestas **independientes**; 3) tamaño suficiente ($n \cdot \hat{p} \geq 10$ y $n \cdot \hat{q} \geq 10$ para usar la normal); 4) población muy grande frente a la muestra, o la corrección por población finita.

ACTIVIDADES

1. Explica, con los datos del estudio de kits, qué son la población, el parámetro, el estimador, la estimación y la variabilidad muestral, y enumera las condiciones de inferencia que se cumplen y cuál habría que comprobar.
2. Si el parámetro fuera p = 0,65 y n = 200, calcula el error estándar y el intervalo en el que caería el 95 % de los p̂. ¿Sería raro obtener 0,60? Calcula su z.
3. Con la hoja de cálculo, obtén la probabilidad de que una variable normal tipificada quede por debajo de 1,645 y entre −2,576 y 2,576. Explica qué relación tienen con los niveles de confianza del 90 % y del 99 %.

10.2 Intervalos de confianza

Un número solo no dice cuánto puede fallar. Un intervalo, sí.

Intervalo de confianza para una proporción. Con muestreo aleatorio simple, población muy grande y las condiciones de tamaño, el intervalo aproximado al nivel de confianza (1 – α) es:

$$\hat{p} \pm z \cdot \sqrt{(\hat{p} \cdot \hat{q} \div n)}$$

donde z es el valor de la normal tipificada que deja en el centro ese nivel (bilateral): 1,645 al 90 %, 1,96 al 95 % y 2,576 al 99 %. El término z · EE es el **error máximo de estimación** o **margen de error** (E), y es la semiamplitud del intervalo. Como p es desconocido, el error estándar se estima con p̂. Un intervalo más ancho es más seguro, pero sirve menos para decidir.

Ejemplo resuelto. Intervalo al 95 % para 120 favorables de 200

Paso	Cálculo	Resultado
Estimación	p̂ = 120 ÷ 200	0,60
Condiciones	n·p̂ = 120; n·q̂ = 80	ambas ≥ 10
Error estándar	√(0,6 · 0,4 ÷ 200) = √0,0012	0,034641
Margen de error	1,96 · 0,034641	0,0679
Intervalo	0,60 ± 0,0679	53,21 % – 66,79 %

En la hoja: =RAIZ(B2*(1-B2)/B3) para el EE y =B2-1,96*B4 y =B2+1,96*B4 para los límites, con p̂ en B2 y n en B3. La semiamplitud es de **6,79 puntos porcentuales**. Si cambias n o el nivel, la hoja recalcula los límites.

Cómo se interpreta: la cobertura del procedimiento. El nivel de confianza no es una propiedad de este intervalo concreto, sino del **método**: si se repitiera muchas veces el muestreo y se calculara cada vez un intervalo así, alrededor del 95 % de esos intervalos contendría el verdadero p. Este intervalo (53,21-66,79 %) lo contiene o no lo contiene: no sabemos cuál de las dos cosas pasa. Por eso no se dice «hay un 95 % de probabilidad de que p esté entre…» ni, menos aún, «el 95 % de las personas está entre…»: el intervalo habla de una proporción de la población, no de respuestas individuales.

Intervalos de muestras repetidas con parámetro fijo (ejemplo ilustrativo, simulado en Python con semilla 2027: 10 muestras de 200 en una población ficticia con p = 58 %)

Muestra	p̂ (%)	Intervalo 95 % (%)	¿Contiene el 58 %?
1	61,5	54,8 – 68,2	sí
2	54,5	47,6 – 61,4	sí
3	64,0	57,3 – 70,7	sí
4	68,5	62,1 – 74,9	no
5	57,5	50,6 – 64,4	sí
6	59,0	52,2 – 65,8	sí
7	59,5	52,7 – 66,3	sí
8	56,5	49,6 – 63,4	sí
9	52,0	45,1 – 58,9	sí
10	63,5	56,8 – 70,2	sí

El parámetro (58 %) no se mueve; los intervalos, sí. Nueve de diez lo atrapan; el de la muestra 4 no. Repetida 1.000 veces, la simulación da 941 intervalos que contienen el 58 %: cerca del 95 % prometido por el método.

Confianza, precisión y tamaño. Con los mismos datos, más confianza exige un intervalo más ancho: al 90 %, 54,30-65,70 % (± 5,70); al 99 %, 51,08-68,92 % (± 8,92). Para estrecharlo sin perder confianza hay que aumentar n: como n está dentro de una raíz, **cuadruplicar la muestra solo divide el margen entre dos** (con 800 encuestados y el mismo 60 %, ± 3,39 puntos). Si se fija n y no se conoce p, el margen máximo se calcula con p = q = 0,5: con n = 200, 1,96 · √(0,25 ÷ 200) = ± 6,93 puntos.

ACTIVIDADES

4. Calcula el intervalo normal para 120 favorables de 200 en población grande y aleatoria simple al 90 % y al 99 %, y explica por qué cambia la amplitud y no cambia el centro.
5. Calcula el intervalo al 95 % si la encuesta hubiera tenido 400 encuestados con 240 favorables. ¿Cuánto se reduce el margen? ¿Por qué no se reduce a la mitad?

10.3 Error muestral y sesgo

Un intervalo estrecho puede estar muy bien calculado y muy mal centrado.

Dos tipos de error. El **error de muestreo** es la diferencia entre p̂ y p que se produce solo por el azar de qué unidades salen. Es aleatorio, se reduce al aumentar n y se puede cuantificar con un intervalo si la selección es probabilística. El **sesgo** (error ajeno al muestreo) es una desviación sistemática: empuja la estimación, de media, hacia el mismo lado, **no se reduce aumentando n** y el intervalo no lo recoge.

Tabla 1 · Fuentes de sesgo en el estudio de kits

Sesgo	Ejemplo en el estudio	¿Lo corrige más n?	Cómo se controla
Cobertura	el listado deja fuera hogares sin correo ni teléfono	no	completar el marco
Selección	contestan voluntarios de las redes	no	selección aleatoria
Medición	pregunta sugestiva	no	pretest; pregunta neutra
No respuesta	no contesta quien tiene menos interés	no	recordatorios; comparar perfiles

Un intervalo estrecho y sesgado. Si Unai publicara una encuesta abierta en redes y contestaran 2.000 voluntarios con un 70 % favorable, la fórmula daría ± 2,0 puntos, un margen sin sentido: supone una muestra aleatoria, y los voluntarios son sobre todo clientes que ya aprecian la tienda. Más respuestas solo estrechan el intervalo alrededor de un valor desplazado.

Ejemplo resuelto. ¿Cuánto puede mover la no respuesta? Supón que para obtener las 200 respuestas se contactó con 250 unidades de la muestra: tasa de respuesta 200 ÷ 250 = **80,0 %**. Si las 50 que no contestaron hubieran sido todas no favorables, la proporción entre los 250 contactados sería 120 ÷ 250 = **48,0 %**; si todas favorables, 170 ÷ 250 = **68,0 %**. Ninguno de los extremos es probable, pero la no respuesta puede desplazar el resultado más que el propio margen de error. Por eso se informa siempre la tasa de respuesta.



ACTIVIDADES

6. Aplica la distinción entre error de muestreo y sesgo al estudio de kits: entrega un ejemplo documentado de cada tipo, una alternativa para controlarlo y la condición que haría cambiar tu elección (por ejemplo, una tasa de respuesta del 40 %). Usa la ficha de la entrega I10.

7. Repite el ejemplo de la no respuesta suponiendo 400 contactos para las mismas 200 respuestas. Calcula la tasa de respuesta y los dos límites, y explica qué conclusión cambia.

10.4 Comparaciones y causalidad

Que dos cosas vayan juntas no significa que una cause la otra.

Comparar grupos. Entre los 200 encuestados había 130 hogares y 70 oficinas:

Tabla 2 · Respuesta favorable por tipo de encuestado

(estudio de kits, mayo de 2027; ficticia)

Tipo	Favorables	Encuestados	% favorable
Hogares	84	130	64,6
Oficinas	36	70	51,4
Total	**120**	**200**	**60,0**

La diferencia es de 13,2 puntos, pero cada grupo tiene su propio margen: al 95 %, oficinas 39,7-63,1 % y hogares 56,4-72,8 %. Antes de atribuirla se revisan la **composición** (¿oficinas grandes o pequeñas?), la **temporalidad** (¿semanas distintas?) y los **factores alternativos** (canal, informante). Una diferencia observada es una **asociación**; para hablar de **causa** hace falta un diseño que controle las alternativas, mejor experimental (grupo de control y asignación aleatoria).

Contraste de hipótesis para una proporción. Nora se ha fijado un umbral: solo prepara el piloto si hay pruebas de que **más de la mitad** de la población encargaría el kit. Un **contraste de hipótesis** decide si los datos son compatibles con una afirmación sobre el parámetro:

1. **Hipótesis.** Nula, H_0: p = 0,50 (no se supera la mitad). Alternativa, H_1: p > 0,50 (contraste unilateral).
2. **Nivel de significación** α = 0,05: la probabilidad de rechazar H_0 siendo cierta (**error de tipo I**) que se acepta. No rechazarla siendo falsa es el **error de tipo II**.
3. **Estadístico.** Suponiendo H_0 cierta, $z = (\hat{p} - p_0) \div \sqrt{(p_0 \cdot q_0 \div n)} = (0{,}60 - 0{,}50) \div \sqrt{(0{,}25 \div 200)} = 0{,}10 \div 0{,}035355 \approx$ **2,828**. Condiciones: 200 · 0,5 = 100 ≥ 10.
4. **p-valor.** Probabilidad de obtener un resultado tan extremo o más si H_0 fuera cierta: P(Z ≥ 2,828) ≈ **0,0023** (en la hoja, =1-DISTR.NORM.ESTAND.N(2,828;VERDADERO)). Equivale a comparar z con el valor crítico unilateral 1,645.
5. **Decisión.** 0,0023 < 0,05: se **rechaza H_0**. Hay evidencia, con muestra aleatoria, de que la proporción de intención favorable supera el 50 %.

Si el umbral fuera el 55 %, z = (0,60 − 0,55) ÷ 0,0352 ≈ 1,42 y el p-valor ≈ 0,0776 > 0,05: **no se rechaza** H_0. Eso no demuestra que p sea 0,55; solo que los datos no bastan para afirmar que lo supera. El p-valor no es la probabilidad de que H_0 sea cierta, y un resultado significativo no dice que la diferencia sea grande ni importante.

Relación entre dos variables cuantitativas. Itxaso ha anotado, en seis semanas de prueba en tienda, las degustaciones ofrecidas y los kits vendidos en tienda (datos ficticios de esta unidad):

Registro 1 · Degustaciones y kits vendidos en tienda

(ficticio)

Semana	Degustaciones (x)	Kits vendidos (y)	$(x - \bar{x})\cdot(y - \bar{y})$
1	10	17	250
2	30	24	15
3	20	23	60
4	50	29	30
5	40	31	20
6	60	38	275
Media / suma	**35**	**27**	**650**

- **Covarianza** muestral: $s_{xy} = \Sigma(x - \bar{x})(y - \bar{y}) \div (n - 1) = 650 \div 5 = 130$. Positiva: las dos variables tienden a subir juntas; su valor depende de las unidades de medida, por eso no sirve para comparar la fuerza de dos relaciones.
- **Coeficiente de correlación lineal de Pearson:** $r = s_{xy} \div (s_x \cdot s_y) = 130 \div (18{,}708 \cdot 7{,}294) \approx$ **0,953**. Va de −1 a 1; cerca de 1, relación lineal positiva fuerte; cerca de 0, sin relación lineal (puede haber otra).
- **Recta de regresión por mínimos cuadrados:** $\hat{y} = a + b\cdot x$, con $b = s_{xy} \div s_x^2 = 650 \div 1.750 \approx$ **0,371** kits por degustación y $a = \bar{y} - b\cdot\bar{x} = 27 - 13 =$ **14** kits. Es la recta que hace mínima la suma de los cuadrados de los residuos $(y - \hat{y})$.
- **Coeficiente de determinación:** $R^2 = r^2 \approx$ **0,908**: el 90,8 % de la variación de los kits vendidos se asocia linealmente con las degustaciones en estas seis semanas.

En la hoja: =COVARIANZA.M, =COEF.DE.CORREL, =ESTIMACION.LINEAL (pendiente y ordenada) y =COEFICIENTE.R2, o un gráfico de dispersión con línea de tendencia, ecuación y R^2. La línea de tendencia también **ajusta curvas** (exponencial, logarítmica, polinómica): se elige por el sentido del fenómeno, no por el R^2 más alto. La recta sirve para predecir **dentro del rango observado** (con 45 degustaciones, 14 + 0,371 · 45 ≈ 30,7 kits); con 150 daría 69,7 kits, una extrapolación sin base.

¿Las degustaciones causan las ventas? No se puede afirmar. Las semanas con más degustaciones son también las últimas (correlación entre semana y degustaciones ≈ 0,886), y los kits eran cada semana más conocidos. Una **regresión múltiple** incluye las dos variables explicativas a la vez, $\hat{y} = a + b_1\cdot x_1 + b_2\cdot x_2$, e interpreta cada coeficiente «manteniendo constante la otra». La herramienta de regresión de la hoja (o =ESTIMACION.LINEAL) da:

Tabla 3 · Salida de regresión múltiple: kits vendidos según degustaciones y semana

(seis observaciones; solo para leerla)

Término	Coeficiente	Lectura
Constante	13,11	kits con 0 degustaciones en la «semana 0»
Degustaciones	0,173	antes 0,371: se mezclaba con la semana
Semana	2,23	kits más por semana, con igual degustación
R^2 · R^2 ajustado	0,978 · 0,964	el ajustado penaliza el número de variables

Con seis semanas y dos variables muy relacionadas, los coeficientes son inestables: la tabla solo enseña a leer la salida, no a decidir con ella. Para saber si las degustaciones venden hace falta un **experimento**: alternar semanas con y sin degustación, al azar, y comparar.

ACTIVIDADES

8. Aplica el contenido al conflicto de la unidad: Unai concluye que «a las oficinas les interesa menos el kit porque es caro». Entrega un ejemplo documentado de por qué la tabla 2 no lo demuestra, una alternativa (qué dato o diseño lo comprobaría) y la condición que haría cambiar tu elección. Usa la ficha de la entrega I10.
9. Contrasta al 5 % si la proporción favorable es distinta de 0,65 (contraste bilateral): plantea H_0 y H_1, calcula z y el p-valor y decide. Comprueba que tu decisión coincide con lo que dice el intervalo al 95 % de 10.2.
10. Con el registro 1 en una hoja de cálculo, calcula covarianza, r, la recta y R^2 con las funciones y con el gráfico de dispersión, y redacta en dos líneas qué se puede afirmar y qué no.

10.5 Conclusiones para decidir

Una buena conclusión responde a la pregunta, dice cuánto, dice con qué seguridad y dice hasta dónde.

Cuatro piezas de una conclusión. 1) **Respuesta al objetivo** del estudio, con sus palabras. 2) **Magnitud** con su incertidumbre: «60 % (IC 95 %: 53,21-66,79 %)». 3) **Límite**: qué no permite afirmar. 4) **Implicación para la decisión**, proporcional a la evidencia: si la evidencia es intermedia, una decisión reversible o condicionada. Si falta una de las cuatro piezas, la conclusión está incompleta.

Intención, probabilidad y compra. Son tres cosas distintas. La **intención declarada** es lo que la persona dice que hará («lo encargaría»). La **probabilidad estimada** es la proporción de la población que declararía esa intención, con su intervalo. La **compra efectiva** es lo que ocurre en caja. Entre la primera y la última hay precio, olvido, horario, competencia y el día a día. Por eso no se calculan «ventas previstas» multiplicando el 60 % por un número de hogares: se mide la conversión real con un piloto. Un 60 % de intención puede acabar en un porcentaje de compra bastante menor, y no se sabe cuánto hasta probarlo.



Ejemplo. Dos conclusiones y una limitación

- **Conclusión 1.** «El 60 % de los encuestados declara que encargaría el kit con entrega programada en el próximo mes; con muestreo aleatorio simple, la proporción en la población se sitúa entre el 53,21 % y el 66,79 % (95 % de confianza) y supera el 50 % (p-valor ≈ 0,0023).»
- **Conclusión 2.** «La intención es mayor en hogares (64,6 %) que en oficinas (51,4 %); la diferencia es descriptiva y las oficinas tienen un margen amplio (n = 70) y habría que confirmarla antes de diseñar una oferta distinta para cada grupo.»
- **Limitación.** «Se mide intención declarada, no compra: el resultado no permite estimar ventas ni la frecuencia de pedido. La no respuesta y la redacción de la pregunta pueden sesgarlo.»

Recomendación proporcional. Con esa evidencia, el equipo no recomienda «lanzar» ni «no lanzar», sino un **piloto de cuatro semanas** con hogares y oficinas, con un indicador de conversión (pedidos reales ÷ contactos) y una fecha de revisión. Si el piloto confirma la demanda, se amplía; si no, se ha aprendido barato.

LA NORMA DICE

Reglamento (UE) 2016/679 (RGPD), de 27 de abril de 2016, considerando 162: el fin estadístico implica que el resultado del tratamiento «no sean datos personales, sino datos agregados», y que ese resultado o los datos personales «no se utilicen para respaldar medidas o decisiones relativas a personas físicas concretas». Las conclusiones son sobre la población; las 200 respuestas no se usan después para enviar ofertas a cada encuestado. Consulta la versión vigente.

ACTIVIDADES

11. Redacta dos conclusiones y una limitación sobre interés declarado y decisión de lanzamiento para el estudio de kits, con magnitud, intervalo y límite, y propón una decisión condicionada con su indicador.
12. Reescribe el titular de Unai («¡El 60 % quiere nuestros kits!») para que sea correcto y siga siendo claro para el equipo de la tienda.

CASO RESUELTO · PASO A PASO

El sesenta por ciento con su intervalo

Caso ficticio. Viernes 28 de mayo de 2027. El equipo de análisis presenta a Nora la inferencia del estudio de kits.

Paso 1. Decisión, datos y alcance. Decisión: si se prepara un piloto del kit con entrega programada. Datos: muestra aleatoria simple de 200 hogares y oficinas (un informante cada uno; sin no respuesta) de una población muy grande; 120 responden «sí, seguro» o «probablemente sí» a la pregunta de intención. Alcance: proporción de **intención declarada** en la población del listado; no ventas.

Paso 2. Procedimiento, con unidades y condiciones. Estimación $\hat{p} = 120 \div 200 = 0{,}60$. Condiciones: $n \cdot \hat{p} = 120 \geq 10$ y $n \cdot \hat{q} = 80 \geq 10$; selección aleatoria; población muy grande. Error estándar: $\sqrt{(0{,}6 \cdot 0{,}4 \div 200)} \approx 0{,}034641$. Margen al 95 % ($z = 1{,}96$, bilateral): $1{,}96 \cdot 0{,}034641 \approx 0{,}0679$, es decir, 6,79 puntos. Intervalo: $0{,}60 \pm 0{,}0679 \rightarrow$ **53,21 %-66,79 %**.

Paso 3. Comprobación e interpretación. Comprobación: el centro del intervalo es $\hat{p}$ ((53,21 + 66,79) ÷ 2 = 60,00) y la semiamplitud es E (66,79 − 60 = 6,79). **Qué permite concluir:** el método, aplicado a muestras así, acierta en el 95 % de los casos; con esta muestra, la intención favorable de la población se estima entre el 53,21 % y el 66,79 %, y el contraste frente al 50 % la sitúa por encima de la mitad. **Qué no permite:** decir que hay un 95 % de probabilidad de que p esté en el intervalo; que el 95 % de las personas esté «dentro»; ni que el 60 % de los hogares vaya a comprar.

RESULTADO

$\hat{p} = 0{,}60$; error $1{,}96 \cdot \sqrt{(0{,}6 \cdot 0{,}4 \div 200)} \approx 0{,}0679$; IC 95 % ≈ 53,21 %-66,79 %. El 95 % es la cobertura del método, no la probabilidad de que el parámetro esté dentro. Decisión propuesta: piloto de cuatro semanas con medición de pedidos reales.

Comprueba que lo has entendido. a) ¿Qué cambiaría si la muestra hubiera sido de voluntarios? b) ¿Cuál sería el intervalo al 99 %? c) ¿Por qué no se calculan ventas con el 60 %?

PRÁCTICA

El servicio de revisión a domicilio de Ruedalia

Datos ficticios. Txema estudia ofrecer una revisión anual de la bicicleta en el domicilio o en la empresa del cliente. Ha encargado una encuesta con **muestreo aleatorio simple de 100 personas usuarias de bicicleta** de una población muy grande, con los mismos supuestos que el caso: **50** responden que contratarían el servicio.

Encargo. 1) Identifica población, parámetro, estimador y estimación. 2) Comprueba las condiciones y calcula el error estándar, el margen y el intervalo al 95 %. 3) Interpreta el intervalo como cobertura del procedimiento. 4) Redacta una conclusión y una limitación para Txema. 5) **Para pensar:** ¿qué cambiaría si las 100 personas fueran las que pasaron por el taller una semana?

Plantilla IC-P10 · Intervalo para una proporción

Campo	Respuesta del alumno
Datos y fuentes con unidad y periodo	
Procedimiento o cálculo	
Resultado y evidencia	
Decisión y limitaciones	

Lista de cotejo

☐ Condiciones comprobadas.
☐ z bilateral del 95 % (1,96).
☐ Intervalo con dos decimales y en porcentaje.
☐ Interpretación como cobertura del método.
☐ Intención separada de contratación real.
☐ Límite por el tipo de muestra.

LECTURA Y DEBATE

El 78 % de Mikel



Relato ficticio. Miércoles 19 de mayo, al salir de una charla de la asociación de comerciantes. Mikel, que lleva una tienda de alimentación en otra ciudad, se acerca a Nora con un café en cada mano. Ha oído que La Despensa estudia los kits y quiere contarle su experiencia.

El año pasado, cuenta, puso unas papeletas junto a la caja: «¿Compraría nuestras cenas preparadas?». En un mes recogió 64 papeletas y 50 decían que sí: **78,1 %**. No dudó: contrató una cocinera media jornada, compró envases para seis meses e imprimió cartelería. A los tres meses vendía menos de una cena al día.

«Lo primero que me preguntó mi gestora», dice Mikel, «fue cuántos tiques habíamos hecho ese mes. Unos 2.400. Me habían contestado 64 personas, el 2,7 % de los tiques, y eran las que se paraban a charlar en la caja: mis clientes más fieles, los que siempre me dicen que sí». Después vino la segunda pregunta: ¿cuántos de esos 50 habían comprado una cena? Nadie lo había apuntado.

Nora le cuenta que su equipo tiene una muestra aleatoria, un intervalo y, aun así, va a hacer un piloto. Mikel se ríe: «Haz el piloto. Y apunta quién compra».

Preguntas. 1) ¿Qué afirmación del caso de Mikel necesitaba una comprobación antes de convertirse en decisión? 2) ¿Qué información habrías pedido primero y por qué habría cambiado la respuesta? 3) Propón la actuación provisional que habría permitido a Mikel aprender sin ocultar la incertidumbre.

APRENDE CON LA IA

Unai prepara el resumen del estudio para el equipo de la tienda y pide a un asistente de IA que le explique el intervalo. Esta es la respuesta:

RESPUESTA DE LA IA

«Un intervalo 95 % significa que el 95 % de las respuestas individuales están dentro. En vuestro caso, 53,21-66,79 % quiere decir que el 95 % de los clientes se sitúa entre esos valores.»

Tu tarea. Localiza el error, contrasta con el procedimiento de la unidad (10.2 y la tabla de muestras repetidas) y escribe una respuesta corregida con su límite: a qué se refiere el 95 %, de qué habla el intervalo y qué no garantiza. Conserva el cálculo que justifica la corrección.

Ojo con la IA. Una explicación que suena sencilla puede cambiar el objeto de la frase: de «el método» a «las personas». Comprueba siempre de qué habla cada porcentaje.

ACTIVIDADES FINALES

Comprueba

13. Resume una distinción que sea decisiva en esta unidad (parámetro y estimación; error de muestreo y sesgo; o intención y compra) y aplícala al caso resuelto.
14. Explica con tus palabras qué significa el p-valor ≈ 0,0023 del contraste frente al 50 % y qué no significa.
15. ¿Por qué R^2 = 0,908 en el registro de degustaciones no demuestra que las degustaciones causen las ventas?

Aplica

16. Repite la práctica de Ruedalia cambiando una condición relevante: confianza del 99 % en lugar del 95 %. Señala qué resultado cambia (amplitud) y cuál se mantiene (estimación) y si cambia la conclusión para Txema.
17. Repite la práctica con 400 personas y 200 favorables. ¿Qué cambia y qué se mantiene? ¿Mejora la conclusión si la muestra fuera de conveniencia?
18. Contrasta con los datos de la práctica (50 de 100) si la proporción supera el 50 % con α = 0,05. Interpreta el resultado.

Investiga

19. Localiza una fuente primaria o documentación técnica pertinente para «conclusiones para decidir»: la nota metodológica de una encuesta del INE o del CIS donde se explique el error de muestreo. Registra fecha de consulta, ámbito, la afirmación que sustenta (cómo expresan la precisión) y una limitación.
20. Busca en la ayuda de tu hoja de cálculo la función de intervalo de confianza que ofrece y comprueba si sirve para proporciones o solo para medias. Anota qué supuestos usa.

TU INVESTIGACIÓN AVANZA · PASO 10 · I10 · INFERENCIA Y CONCLUSIÓN

Con los datos ficticios o propios autorizados de tu investigación: 1) población, parámetro y estimador; 2) condiciones de la inferencia comprobadas (selección, independencia, tamaño); 3) intervalo con su cálculo en la hoja y su interpretación como cobertura del método; 4) tabla de fuentes de sesgo con su control; 5) si comparas grupos o variables, la asociación sin atribuir causa, y un contraste si hay un umbral de decisión; 6) dos conclusiones y una limitación; 7) decisión condicionada, responsable y fecha de revisión. Modelo: 120 de 200 (53,21-66,79 %).
Comprueba tu paso: ningún margen de error para una muestra no aleatoria; intención separada de compra; fórmulas visibles.
☐ Hecho ☐ Revisado ☐ Guardado

OJO PROFESIONAL

Documento: resumen de resultados de la encuesta en redes

Documento ficticio, preparado por Unai el lunes 17 de mayo de 2027 para la reunión de La Despensa de Nora, S.L. Contiene **cinco errores**.

N.º	Apartado	Contenido
1	Método	Encuesta abierta en las redes sociales de la tienda, del lunes 10 al viernes 14 de mayo de 2027; 312 respuestas
2	Resultado	218 de 312 responden «sí» o «probablemente sí» (69,9 %)
3	Precisión	«Margen de error de ± 5,1 puntos con un 95 % de confianza»
4	Interpretación	«El 95 % de los clientes del barrio está entre el 64,8 % y el 75,0 %»
5	Ventas	«Si el barrio tiene unos 6.000 hogares, venderemos 4.200 kits al mes»
6	Contraste	«p-valor inferior a 0,001: hay más de un 99,9 % de probabilidad de que la mayoría quiera el kit»
7	Causa	«Las semanas con más degustaciones vendimos más kits: las degustaciones causan las ventas»

Encargo. Anota las cinco correcciones justificadas (línea, qué falla y cómo debe quedar) y reescribe el resumen para que se pueda utilizar: qué describe, de quién, con qué límites y qué decisión permite. **Para pensar:** ¿qué dos líneas son correctas tal como están y por qué la segunda solo vale para los que respondieron?

N.º	Línea	Por qué es un error	Corrección
1			
2			
3			
4			
5			

AUTOEVALUACIÓN

1. En el estudio de kits, $\hat{p}$ = 0,60 es...
a) el parámetro, porque sale de una muestra aleatoria.
b) la estimación que da el estimador con esta muestra; el parámetro p es desconocido.
c) la probabilidad de que cada hogar compre.

2. Para 120 favorables de 200 (m.a.s., población grande), el intervalo al 95 % es...
a) 53,21 %-66,79 %.
b) 56,54 %-63,46 %.
c) 51,08 %-68,92 %.

3. Una encuesta de 2.000 voluntarios da 70 % ± 2,0 puntos. Lo correcto es decir que...
a) es muy precisa, porque el margen es pequeño.
b) el margen no es válido: la selección voluntaria es un sesgo que el tamaño no corrige.
c) habría que llegar a 4.000 respuestas.

4. Hogares 64,6 % y oficinas 51,4 % favorables. Se puede afirmar que...
a) ser oficina causa menos interés.
b) hay una asociación descriptiva que hay que revisar por composición, momento y otros factores.
c) la diferencia es significativa porque supera 10 puntos.

5. Una conclusión útil para Nora es...
a) «El 60 % comprará el kit».
b) «Venderemos 3.600 kits si hay 6.000 hogares».
c) «Intención favorable del 60 % (IC 95 %: 53,21-66,79 %); no mide compra: piloto con pedidos reales».

6. El margen de error al 95 % del caso (120 de 200) es...
a) 0,0346.
b) 0,0679.
c) 0,0892.

7. En la práctica de Ruedalia (50 de 100, mismos supuestos), el intervalo al 95 % es...
a) 40,2 %-59,8 %.
b) 45,1 %-54,9 %.
c) 37,1 %-62,9 %.

8. «Un intervalo 95 % significa que el 95 % de las respuestas individuales están dentro» es falso porque...
a) debería ser el 99 %.
b) el 95 % se refiere a la cobertura del método que estima el parámetro, no a las respuestas.
c) solo vale para muestras grandes.

9. En el resumen de Unai, el «± 5,1 puntos» es un error porque...
a) está mal calculado.
b) se aplica una fórmula de muestreo aleatorio a respuestas voluntarias.
c) debería darse con tres decimales.

10. En tu entrega I10, un dato adicional que te haría revisar la conclusión sería...
a) la tasa de respuesta y el perfil de quienes no contestaron.
b) un gráfico más llamativo.
c) el resultado de otro estudio del encargo.

Ya sé...

- ☐ distinguir estimador, intervalo, error y sesgo, asociación y causa, conclusión.
- ☐ ejecutar el procedimiento del caso: intervalo de 120 de 200 y su contraste frente al 50 %.
- ☐ transferirlo a otro negocio, como la revisión a domicilio de Ruedalia.
- ☐ detectar un margen de error aplicado a respuestas voluntarias.
- ☐ entregar evidencia revisable: hoja con fórmulas, condiciones y límites.



UNIDAD 11

Base de datos relacional

Bloque 4 · Análisis y comunicación de resultados

EN ESTA UNIDAD APRENDERÁS A…

- Identificar entidades, tablas, registros, campos, claves y relaciones, y saber qué representa cada fila.
- Diseñar una base de datos relacional con claves primarias y foráneas, y comprobar su integridad.
- Ejecutar consultas de selección, agregación, unión y acción, y contar personas, encuestas y respuestas sin confundirlas.
- Construir formularios con validación y probarlos con datos válidos e inválidos.
- Proteger, copiar, restaurar y mantener la base, y crear informes, etiquetas y exportaciones.

SITUACIÓN DE PARTIDA

Lunes 31 de mayo de 2027. Las encuestas del **Estudio de kits 2027** están en hojas de cálculo sueltas, y Nora quiere una sola base para el informe de junio. Unai lo tiene claro: «Una hoja grande, una fila por respuesta, y en cada fila el nombre, la dirección y el teléfono de quien contesta».

- Adrián, que pasa a limpio los cuestionarios, avisa de que quien conteste diez preguntas saldría diez veces con su dirección.
- Olvido pregunta quién va a ver esos teléfonos.
- Nora solo quiere saber cuántas personas han contestado de verdad.

Nora te encarga la base **EK-2027/BD-v1.0**: se prueba con las encuestas ficticias E01 (martes 1/06) y E02 (miércoles 2/06) y se da por comprobada el viernes 11/06.

PARA EMPEZAR

1. ¿Qué tiene que decidir o comprobar Nora: cómo guardar las respuestas, cuántas personas han contestado o las dos cosas?
2. ¿Qué sabemos de verdad (cuestionarios, fechas, respuestas) y qué suponemos al decir que cada fila es una persona?
3. ¿Qué prueba con la base permitiría revisar la primera respuesta de Unai antes de cargar los datos?
4. ¿Qué dato personal hace falta para analizar una encuesta y cuál sobra?



Entidades y claves

11.1



Diseño e integridad

11.2



Consultas

11.3



Formularios

11.4



Seguridad e informes

11.5

11.1 Entidades, claves y relaciones

Antes de contar nada, pregunta qué representa cada fila.

De la hoja de cálculo a la base de datos. Una hoja de cálculo con una fila por cuestionario sirve mientras hay un solo tipo de dato. En el estudio de kits hay tres: las **personas** (hogares u oficinas) que participan, las **encuestas** que cada una contesta en una fecha y por un canal, y las **respuestas** a cada pregunta. Mezcladas en una hoja, el mismo dato se repite y nadie sabe ya qué representa una fila. Una **base de datos relacional** guarda cada dato una sola vez, en tablas que se enlazan por claves. La gestiona un **gestor de bases de datos** (LibreOffice Base u otro), que entiende el lenguaje de consultas **SQL** estándar.

Tabla 1 · Elementos de una base de datos en el estudio de kits

Elemento	Qué es	En el estudio de kits
Entidad y tabla	Aquello de lo que se guardan datos	Persona → Personas
Registro y campo	Fila (un caso); columna (un dato)	E01; su fecha
Clave primaria	Identifica cada registro sin repetirse	id_encuesta
Clave foránea	Apunta a la clave primaria de otra tabla	id_persona en Encuestas
Relación	Enlace entre tablas por una clave	Persona 1 → N encuestas
Consulta · formulario · informe	Pregunta · pantalla de entrada · salida	Respuestas por encuesta

La granularidad: qué es una fila. La **granularidad** es el nivel de detalle de cada registro. En Personas, una fila es un hogar u oficina que participa; en Encuestas, una vez que esa persona contesta, en una fecha y por un canal; en Respuestas, la respuesta a una pregunta de una encuesta. Las tres tablas **no se cuentan del mismo modo**: diez filas de Respuestas pueden ser una sola persona. Si al preguntar qué representa una fila la respuesta es «depende», la tabla está mal diseñada.

Ejemplo. Tres hojas, tres granularidades. La hoja de la tarjeta de cliente de 2026 tenía una fila por titular; la de la encuesta del reparto, una por cuestionario; la de Unai, una por respuesta. Sumar sus filas como «clientes estudiados» contaría varias veces a la misma persona: hay que enlazar por un identificador común y contar valores distintos.

Las relaciones. Una persona puede contestar más de una encuesta (si el estudio la repite en septiembre) y cada encuesta tiene varias respuestas: son relaciones **uno a varios** (1:N). Las de **varios a varios** (N:M) se resuelven con una tabla intermedia: personas y kits probados se enlazarían con una tabla Pruebas.

El catálogo de datos. Un dato que no se sabe dónde está no se puede presentar. El equipo documenta cada tabla en el **diccionario de datos** del estudio (EK-2027/DD-v1.0): archivo, tabla, campo, tipo, valores, clave y versión. La base se guarda como 2027-06-11_estudio_kits_base_v1.0, y los datos brutos, aparte y sin tocar.

Tabla 2 · Catálogo de la tabla Encuestas (EK-2027/BD-v1.0; extracto)

Campo	Tipo y valores permitidos	Clave	Qué representa
id_encuesta	texto, E01, E02…	primaria	Una encuesta contestada
id_persona	texto, P01, P02…	foránea → Personas	Quién la contesta
canal	lista: presencial, web, telefono	—	Modo de recogida

ACTIVIDADES

1. Distingue en el caso una tabla, un registro, un campo, una clave primaria, una clave foránea, una relación, una consulta, un formulario y un informe. Para cada uno, escribe un ejemplo distinto de los de la tabla 1.
2. Documenta en un catálogo como el de la tabla 2 las tablas Personas y Respuestas: archivo, tabla, campo, tipo, unidad o valores, clave y versión. Indica qué representa una fila de cada tabla.

11.2 Diseño e integridad

Normalizar es guardar cada dato una sola vez sin perder ninguna relación.

El problema de la hoja única. La propuesta de Unai es una **tabla plana**: todos los datos en una hoja, una fila por respuesta:

Tabla 3 · Hoja plana propuesta por Unai

(datos ficticios; nombre, dirección y teléfono omitidos)

Respuesta	Persona, tipo y zona	Encuesta, fecha y canal	Pregunta y valor
R01	P01 · hogar · Norte	E01 · 2027-06-01 · presencial	P1_interes · 4
R02	P01 · hogar · Norte	E01 · 2027-06-01 · presencial	P2_precio · 3
R03	P02 · oficina · Centro	E02 · 2027-06-02 · web	P1_interes · 5

Con solo tres respuestas, P01 y E01 ya salen dos veces. Esa **redundancia** produce tres fallos: **de actualización** (si el hogar P01 cambia de zona, hay que corregir todas sus filas y olvidar una deja dos zonas); **de inserción** (no se puede dar de alta una persona que aún no ha contestado) y **de borrado** (al borrar la única respuesta de una encuesta se pierde la encuesta). Y repetir nombre, dirección y teléfono multiplica los datos personales: con dos personas y diez preguntas, la hoja plana tendría 20 filas con sus datos personales; la base normalizada, 2 filas con ellos (tabla Contactos) y 20 sin ellos (Respuestas).

Normalizar. Normalizar es repartir los datos en tablas para guardarlos una vez. Se comprueba con las tres primeras **formas normales**: 1) cada campo guarda un valor y no hay columnas repetidas (P1, P2, P3…); 2) cada campo depende de toda la clave, no de una parte (con clave id_encuesta + pregunta, el texto dependería solo de pregunta); 3) ningún campo depende de otro no clave: la zona, de la persona; el canal, de la encuesta.

Tabla 4 · Diagrama relacional de EK-2027/BD-v1.0

Tabla (una fila es…)	Clave primaria	Clave foránea	Otros campos
Personas (hogar u oficina)	id_persona	—	tipo, zona
Encuestas (una contestada)	id_encuesta	id_persona	fecha, canal
Respuestas (a una pregunta)	id_respuesta	id_encuesta	pregunta, valor

Relaciones: Personas 1 → N Encuestas → N Respuestas. La clave foránea va siempre en la tabla del lado «varios»: id_persona en Encuestas, no una lista de encuestas en Personas. Respuestas no lleva id_persona: la persona se obtiene por la encuesta. En SQLite (en un gestor gráfico: diseñador de tablas y relaciones):

- CREATE TABLE Personas (id_persona TEXT NOT NULL PRIMARY KEY, tipo TEXT NOT NULL CHECK (tipo IN ('hogar', 'oficina')), zona TEXT NOT NULL);
- CREATE TABLE Encuestas (id_encuesta TEXT NOT NULL PRIMARY KEY, id_persona TEXT NOT NULL REFERENCES Personas (id_persona), fecha TEXT NOT NULL, canal TEXT NOT NULL CHECK (canal IN ('presencial', 'web', 'telefono')));
- CREATE TABLE Respuestas (id_respuesta TEXT NOT NULL PRIMARY KEY, id_encuesta TEXT NOT NULL REFERENCES Encuestas (id_encuesta), pregunta TEXT NOT NULL, valor INTEGER NOT NULL CHECK (valor BETWEEN 1 AND 5));

Integridad. La **integridad de entidad** exige una clave primaria única y no vacía. La **integridad referencial** exige que toda clave foránea apunte a un registro existente: no puede haber una respuesta **huérfana** (de una encuesta inexistente) ni borrarse una persona con encuestas. CHECK y NOT NULL son **reglas de dominio**. Algunos gestores traen las claves foráneas desactivadas (en SQLite, PRAGMA foreign_keys = ON): la base acepta huérfanas sin avisar.

Ejemplo resuelto. Prueba de integridad del martes 8 de junio. Con los datos de prueba cargados, el equipo ensaya cinco entradas:

Tabla 5 · Registro de pruebas de integridad (EK-2027/BD-v1.0)

Prueba	Entrada	Esperado	Obtenido
Alta válida	R04 → E02, P2_precio, 4	aceptada	aceptada (luego se borra)
Respuesta huérfana	R99 → E09 (no existe)	rechazada	rechazada
Valor fuera de rango	R98 → E01, P1_interes, 7	rechazada	rechazada
Clave repetida	otra R01	rechazada	rechazada
Borrar con dependientes	borrar P01, que tiene E01	rechazada	rechazada

LA NORMA DICE

Reglamento (UE) 2016/679 (RGPD), de 27 de abril de 2016, art. 5.1.c): los datos serán «adecuados, pertinentes y limitados a lo necesario» (minimización). Art. 25.1: medidas apropiadas desde el diseño, «como la seudonimización». Consulta la versión vigente.

Privacidad desde el diseño. Para analizar basta con el tipo y la zona: la base no guarda nombre ni dirección. Si hay que volver a llamar a alguien, el teléfono va en una tabla aparte de contactos, con acceso solo para quien llama y borrado al terminar el campo. Eso es **seudonimizar**: el análisis trabaja con P01 y P02, y solo quien tiene los contactos sabe quién es quién. Un dato seudonimizado sigue siendo personal; uno anonimizado ya no identifica a nadie.

ACTIVIDADES

3. Construye las tres tablas del caso con sus claves primarias y foráneas en un gestor relacional (LibreOffice Base u otro), carga P01-P02, E01-E02 y R01-R03, y repite las cinco pruebas de la tabla 5. Guarda el registro de pruebas con fecha y versión.
4. Con la hoja de la tabla 3, pon un ejemplo concreto de fallo de actualización, otro de inserción y otro de borrado.
5. Unai quiere guardar el teléfono «por si hay que volver a llamar». Propón dónde se guarda, quién lo ve y cuándo se borra, y explica qué cambiaría si no se fuera a volver a llamar a nadie.

11.3 Consultas y búsquedas

Una consulta es una pregunta: antes de ejecutarla, escribe qué resultado esperas.

Tipos de consulta. Las de **selección** devuelven filas que cumplen una condición (WHERE) sin cambiar nada. Las de **agregación** resumen con COUNT, AVG, MIN o MAX, y con GROUP BY, por grupos. Las de **unión** (JOIN) combinan tablas por sus claves. Las de **acción** cambian datos: INSERT, UPDATE y DELETE. Las búsquedas avanzadas combinan filtros (AND, OR), texto parcial (LIKE 'E0%') o rangos de fechas (BETWEEN). En un gestor gráfico se construyen con el asistente de consultas y se **guardan con nombre** (q_respuestas_por_encuesta) para repetirlas con más datos.

Ejemplos: **selección**, SELECT id_encuesta FROM Encuestas WHERE canal = 'web'; → E02. **Media**, SELECT AVG(valor) FROM Respuestas WHERE pregunta = 'P1_interes'; → 4,5 (2 respuestas; SQLite, otros truncan). **Agregación**, SELECT id_encuesta, COUNT(*) FROM Respuestas GROUP BY id_encuesta; → E01: 2 · E02: 1.

La unión de las tres tablas. Para saber qué respondió cada persona se unen las tres tablas por sus claves (Personas p JOIN Encuestas e ON e.id_persona = p.id_persona JOIN Respuestas r ON r.id_encuesta = e.id_encuesta).

Tabla 6 · Resultado del JOIN (una fila por respuesta)

Persona	Encuesta	Respuesta	Pregunta y valor
P01	E01	R01	P1_interes · 4
P01	E01	R02	P2_precio · 3
P02	E02	R03	P1_interes · 5

El JOIN hereda la granularidad de la tabla más detallada: **cada fila es una respuesta**, no una persona. Para contar, hay que decir qué se cuenta:

SELECT COUNT(*), COUNT(DISTINCT e.id_encuesta), COUNT(DISTINCT p.id_persona) FROM Personas p JOIN Encuestas e ON e.id_persona = p.id_persona JOIN Respuestas r ON r.id_encuesta = e.id_encuesta;

Resultado: **3** respuestas, **2** encuestas y **2** personas. COUNT(*) cuenta filas; COUNT(DISTINCT campo), valores distintos de ese campo.



Búsquedas avanzadas y de control. Una búsqueda combina condiciones: SELECT id_encuesta FROM Encuestas WHERE canal = 'web' AND fecha BETWEEN '2027-06-01' AND '2027-06-04'; → E02. Una **consulta de control** busca lo que no debería existir, como encuestas sin respuestas: SELECT e.id_encuesta FROM Encuestas e LEFT JOIN Respuestas r ON r.id_encuesta = e.id_encuesta WHERE r.id_respuesta IS NULL; → ninguna fila. El LEFT JOIN conserva las encuestas sin pareja.

Consultas de acción, siempre en una copia. Una acción no se deshace con un clic. Cinco pasos: 1) **copia** de la base con fecha; 2) selección previa con la misma condición, para ver qué filas cambian; 3) acción en la copia; 4) comprobación después; 5) anotación en el registro de cambios. Prueba del 8 de junio, en la copia: web antes = 1; UPDATE Encuestas SET canal = 'telefono' WHERE id_encuesta = 'E02'; después = 0. El original sigue con E02 en web.

CLAVE · CONTAR CON DISTINCT

Tras un JOIN, cuenta personas con COUNT(DISTINCT id_persona) y escribe siempre qué has contado.

ACTIVIDADES

6. Crea una consulta de selección, una de agregación y una de acción sobre una copia, cada una con su finalidad, el resultado esperado y el obtenido, y el recuento antes y después de la acción.
7. El jueves 3 de junio P02 contesta por la web una segunda encuesta, E03, con una respuesta, R04 (P1_interes, 3). Predice el JOIN y las tres cuentas con DISTINCT, y compruébalo.
8. Une las tres tablas, filtra P1_interes y calcula la media por tipo de participante con GROUP BY. Con la tabla 6, ¿qué base tiene cada media y qué permite decir?

11.4 Formularios y validación

Un formulario no está terminado cuando se ve bien, sino cuando rechaza lo que debe rechazar.

Qué es un formulario. Un **formulario** es una pantalla vinculada a una tabla para dar de alta, consultar o editar registros sin tocarla directamente. Adrián pasa a la base los cuestionarios en papel de la tienda; el formulario le evita teclear códigos y le impide guardar datos imposibles. Los gestores tienen un asistente que lo crea a partir de la tabla; después se ajusta a mano.



Los controles. El formulario de alta de respuestas tiene un **selector de encuesta** (lista de las encuestas que existen: no se puede escribir E09), una **lista de preguntas** con los códigos del cuestionario, un **control de valores permitidos** (botones de 1 a 5; «no sabe» y «no procede» no son valores: se documentan aparte en el diccionario) y **campos obligatorios** con mensaje claro. Las encuestas se dan de alta con un **subformulario**: arriba, persona, fecha y canal; debajo, sus respuestas.

Dos niveles de validación. Los datos importados de un formulario en línea no pasan por el formulario del gestor. Por eso las mismas reglas están también en la tabla (CHECK, NOT NULL y clave foránea): si el formulario falla, la tabla rechaza el dato.

Probar el formulario. Una imagen no demuestra que funcione. Se prueba, en una copia descartable, con un **plan de pruebas**: una entrada válida, una inválida y una referencia inexistente, con el resultado esperado y el obtenido.

Tabla 7 · Plan de pruebas del formulario (09/06/2027, v1.0)

Prueba	Entrada	Esperado	Obtenido
Alta válida	E02 · P2_precio · 4	se guarda	se guarda
Dato inválido	E01 · P1_interes · 7	no deja guardar	no deja guardar
Referencia inexistente	encuesta E09	no está en la lista	no está en la lista
Campo vacío	E01 · sin pregunta	mensaje y no guarda	mensaje y no guarda

Formularios en línea. Uno en línea (Google Forms u otro) guarda una hoja plana: se pasa a Encuestas y Respuestas comprobando los recuentos, y la hoja original se guarda sin tocar.

ACTIVIDADES

9. Construye un formulario de alta de respuestas con validación, selector de encuesta y control de valores permitidos. Pruébalo con una alta válida, un dato inválido y una referencia inexistente, y entrega el plan de pruebas.

10. En la hoja de Unai, el canal se escribía a mano y aparecen «web», «Web», «WEB» y «web » (con espacio final). ¿Cuántos canales contaría una agrupación en SQLite? Propón el control del formulario y la regla de la tabla que lo evitan.

11.5 Seguridad, mantenimiento e informes

Cada salida tiene un destinatario, los datos que necesita y una versión que se puede rastrear.

Permisos y roles. El equipo define **roles** con el principio de **mínimo privilegio**: cada persona tiene solo los permisos que su tarea exige.

Tabla 8 · Roles de EK-2027/BD-v1.0

Rol	Quién	Puede	No puede
Lectura	Nora, Olvido	Ver informes	Editar nada
Edición	Adrián, Unai	Altas con el formulario	Cambiar tablas, borrar
Administración	Equipo de análisis	Estructura, copias, permisos	—

En un gestor con servidor se asignan con GRANT y REVOKE (GRANT SELECT ON v_informe TO lectura; es una vista de informe); en una de escritorio, con permisos de carpeta e informes de solo lectura. Se **prueban**: con el rol de lectura se intenta editar y el sistema debe impedirlo; con el de edición se da de alta una respuesta y se intenta cambiar una tabla. Cada prueba se anota con fecha y los permisos se revisan cuando alguien entra o sale.

Copia y restauración. Copia al cierre de cada día de campo y antes de cada cambio de estructura, con fecha en el nombre y en otro soporte. No está comprobada hasta que se **restaura** en otra carpeta y las consultas de control dan los mismos recuentos (3 respuestas, 2 encuestas y 2 personas). Se conservan las copias de los últimos cinco días de campo y la de cada versión cerrada.

LA NORMA DICE

RGPD, art. 5.1.f): los datos se tratarán con «una seguridad adecuada» («integridad y confidencialidad»). Art. 32.1.c): medidas que incluyan «la capacidad de restaurar la disponibilidad y el acceso a los datos personales de forma rápida en caso de incidente físico o técnico». Ley Orgánica 3/2018 (LOPDGDD), de 5 de diciembre, art. 5: todas las personas que intervengan en el tratamiento están sujetas al deber de confidencialidad, también al terminar su relación. Consulta la versión vigente.

Mantenimiento. Un **registro de cambios** con versión, fecha, motivo y responsable (v0.9 → v1.0: «se activa la clave foránea de Respuestas»), consultas de control periódicas y un **plazo de conservación**: los contactos se borran al terminar el campo.

Confidencialidad de los datos de investigación. El **Código Internacional ICC/ESOMAR** (edición de 2025), código de autorregulación de la profesión, no una ley, pide una investigación legal, honesta y transparente, que no perjudique a quien participa y proteja sus datos. Tres reglas del equipo: los datos del estudio solo sirven al estudio (enviar ofertas a cada encuestado ya no es investigar, sino vender); los resultados se publican **agregados**; y quien toca la base guarda confidencialidad, también al terminar. Consulta la versión vigente del código.

Informes, etiquetas y exportaciones. Un **informe** es una salida vinculada a una consulta guardada: si cambian los datos, se regenera con fecha y versión. Las **etiquetas** de campo llevan identificadores ficticios (E03, E04…), fecha prevista y canal, sin nombres. Cada **exportación** se diseña para su destinatario:

Tabla 9 · Salidas de EK-2027/BD-v1.0

Salida	Destinatario	Campos	Formato
Resumen de participación	Nora (dirección)	totales y media de P1 con su base	PDF con versión
Vista de análisis	Equipo de análisis	id_encuesta, tipo, zona, pregunta, valor	CSV UTF-8 con diccionario
Etiquetas de campo	Adrián	id_encuesta, fecha prevista, canal	Hoja de etiquetas

Nora decide con totales, sin identificadores; el análisis usa códigos, nunca contactos, y el diccionario, sin el que «P1_interes = 4» no significa nada. El CSV se exporta en UTF-8 y con la fecha y la versión en el nombre, para saber de qué base sale cada cifra del informe.

EN TU SECTOR

Marketing y publicidad: la base de una campaña enlaza piezas, soportes e impactos; contar impactos no es contar personas alcanzadas. **Gestión de ventas y espacios comerciales:** la base de visitas enlaza clientes, visitas y pedidos; diez visitas pueden ser un solo cliente.

ACTIVIDADES

11. Prueba en tu base un rol de lectura y uno de edición (o simúlalos con permisos de carpeta y un informe de solo lectura), haz una copia de seguridad y restáurala en otra carpeta. Entrega las pruebas con los recuentos antes y después.
12. Exporta una vista para dirección y otra para análisis. Justifica qué campos lleva cada una, cuáles no lleva y por qué formato has optado.
13. Genera un informe de resultados vinculado a una consulta guardada y una hoja de etiquetas con identificadores ficticios (E03 a E10) para el material de campo. Comprueba que el informe cambia al añadir una respuesta.

CASO RESUELTO · PASO A PASO

La base del estudio de kits

Caso ficticio. El viernes 11 de junio de 2027 el equipo de análisis presenta a Nora la base EK-2027/BD-v1.0, probada con dos encuestas ficticias de prueba.

Paso 1. Decisión, datos y alcance. Nora decide si la base sustituye a las hojas sueltas para el resto del campo. Datos: Personas P01 (hogar, zona Norte) y P02 (oficina, zona Centro); Encuestas E01 → P01 (martes 01/06, presencial) y E02 → P02 (miércoles 02/06, web); Respuestas R01 → E01 (P1_interes, 4), R02 → E01 (P2_precio, 3) y R03 → E02 (P1_interes, 5). Alcance: comprobar el diseño y las consultas, no medir el interés del barrio.

Paso 2. Procedimiento. 1) Claves: id_persona, id_encuesta e id_respuesta, primarias; id_persona en Encuestas e id_encuesta en Respuestas, foráneas, con la integridad referencial activada. 2) Carga de los siete registros (2 + 2 + 3). 3) Consulta por encuesta: SELECT id_encuesta, COUNT(*) FROM Respuestas GROUP BY id_encuesta; → E01: 2 respuestas; E02: 1 respuesta. 4) JOIN de las tres tablas: 3 filas, una por respuesta. 5) Recuentos: filas del JOIN = 3 respuestas; COUNT(DISTINCT id_encuesta) = 2 encuestas; COUNT(DISTINCT id_persona) = 2 personas.

Paso 3. Comprobación e interpretación. La suma por encuesta (2 + 1) coincide con las 3 filas del JOIN. La respuesta huérfana R99 → E09 se rechaza; también el valor 7, la R01 repetida y el borrado de P01. La copia restaurada da los mismos recuentos. Permite concluir que la base guarda cada dato una vez, enlaza bien las tres tablas y cuenta bien. **No** permite concluir nada sobre el interés por los kits: una media de 4,5 sobre 2 respuestas de prueba no describe a nadie más.

RESULTADO

Dos personas, dos encuestas y tres respuestas. Un JOIN devuelve una fila por respuesta: para contar personas hace falta COUNT(DISTINCT id_persona). La base rechaza la respuesta huérfana. Nora aprueba EK-2027/BD-v1.0 para el resto del campo, con la tabla de contactos separada y de acceso restringido, y la revisión de recuentos cada viernes.

Tabla 10 · Recuentos de control de la base (viernes 11/06/2027)

Qué se cuenta	Consulta	Resultado
Respuestas	COUNT(*) del JOIN	3
Encuestas	COUNT(DISTINCT id_encuesta)	2
Personas	COUNT(DISTINCT id_persona)	2
Huérfanas	respuestas sin encuesta	0

Comprueba que lo has entendido. a) Si la consulta por encuesta diera E01: 2 y E02: 2, ¿qué recuento de control fallaría? b) ¿Por qué la media de 4,5 no sirve para el informe de junio? c) ¿Qué pasaría con la prueba de R99 si la integridad referencial no estuviera activada?

PRÁCTICA

La base de satisfacción de Ruedalia

Caso ficticio. Txema quiere saber qué opinan los clientes del servicio a empresas y ha empezado una base **con la misma estructura** que la de La Despensa (CHECK de tipo: particular o empresa): Personas P01 (particular, polígono) y P02 (empresa, polígono); Encuestas E01 → P01 (lunes 07/06/2027, presencial) y E02 → P02 (martes 08/06/2027, web); Respuestas R01 → E01 (P1_satisfaccion, 4), R02 → E01 (P2_plazo, 3) y R03 → E02 (P1_satisfaccion, 2). El jueves 10 de junio Ousmane hace por teléfono una segunda encuesta a P01.

Encargo. 1) En un gestor relacional, crea las tres tablas con sus claves y la integridad referencial activada, y carga los datos. 2) Añade la encuesta **E03 → P01** (10/06/2027, teléfono) y dos respuestas de E03: **R04** (P1_satisfaccion, 5) y **R05** (P2_plazo, 4). 3) Ejecuta la consulta por encuesta y el JOIN con COUNT(*) y COUNT(DISTINCT …). 4) Prueba una respuesta huérfana y un valor fuera de rango. 5) Crea un formulario de alta de respuestas y un informe con los recuentos para Txema.

Plantilla IC-P11 · Base de Ruedalia

Campo	Tu respuesta
Datos y fuentes con unidad y periodo	
Procedimiento o cálculo	
Resultado y evidencia	
Decisión y limitaciones	

Para pensar. Txema dice: «Ya tenemos tres clientes opinando». ¿Qué le contestas, con qué consulta y qué pruebas le enseñas?

Lista de cotejo

☐ Tres tablas con claves primarias y foráneas.
☐ E03 enlazada a P01.
☐ Recuentos de respuestas, encuestas y personas con su consulta.
☐ Huérfana y valor fuera de rango rechazados.
☐ Formulario e informe funcionando, no solo dibujados.

LECTURA Y DEBATE

La tarjeta que contaba dos veces



Relato ficticio. La responsable de fidelización de una pequeña cadena de tiendas de deportes tenía el teléfono en la oreja y la tarjeta de una socia sobre la mesa. La socia estaba molesta: le habían llegado tres cartas de la campaña de primavera, dos a una dirección en la que ya no vivía.

La campaña había salido de la hoja de socios, que en realidad era la hoja de compras: una fila por compra, y en cada fila el nombre, la dirección y el teléfono que la cajera tecleaba ese día. Con esa hoja había informado a la dirección de que la tarjeta tenía «1.200 socios activos», y con esa cifra se habían encargado 1.200 cartas. Nadie había preguntado qué era una fila.

Esa tarde contó los números de tarjeta distintos: 430. Las 1.200 filas eran compras. Además, 26 socios tenían dos direcciones distintas, según la tienda y el día. La primera decisión, mandar una carta por fila, había costado casi tres veces lo necesario y había molestado justo a las socias más fieles, las que más compraban.

No borró nada. Guardó la hoja como estaba, separó a los socios en una tabla con una fila por tarjeta y las compras en otra, enlazadas por el número de tarjeta, y escribió a la dirección: «La cifra correcta es 430 socios con al menos una compra. Para la próxima campaña propongo confirmar la dirección en caja antes de enviar nada».

Preguntas. 1) ¿Qué afirmación del caso necesitaba una comprobación antes de convertirse en decisión? 2) ¿Qué información pedirías primero y por qué cambiaría la respuesta? 3) Propón una actuación provisional para la próxima campaña que permita aprender sin ocultar la incertidumbre sobre las direcciones.

APRENDE CON LA IA

Unai pega en un asistente de IA el resultado del JOIN de la tabla 6 y le pregunta cuántas personas han participado. Esta es la respuesta:

RESPUESTA DE LA IA

«Has unido Personas, Encuestas y Respuestas y la consulta devuelve tres filas. Tres filas tras unir tablas demuestran que hay tres entrevistados. Puedes informar a Nora de que han participado tres personas en la prueba de la base.»

Tu tarea. Localiza el error, contrástalo con el procedimiento de la unidad (qué representa una fila del JOIN y cómo se cuentan las personas) y escribe una respuesta corregida que diga también su límite. Conserva la consulta y su resultado como prueba de la corrección.

Ojo con la IA. Un asistente no ve tu base: solo lee lo que le pegas. No le pegues datos de contacto de personas reales.

ACTIVIDADES FINALES

Comprueba

14. Resume una distinción decisiva de esta unidad (por ejemplo, fila del JOIN frente a persona, o validación del formulario frente a regla de la tabla) y aplícala al caso resuelto.
15. Explica la diferencia entre clave primaria y clave foránea con un ejemplo del caso, y qué impide cada una.
16. ¿Por qué una consulta de acción se ejecuta primero en una copia? Enumera los cinco pasos del procedimiento.
17. ¿Cuándo está comprobada una copia de seguridad?

Aplica

18. Repite la práctica de Ruedalia cambiando una condición relevante: la encuesta E03 la contesta P02 en lugar de P01, con las mismas dos respuestas. Señala qué resultados cambian (respuestas por encuesta, encuestas por persona) y cuáles se mantienen.
19. Una oficina del estudio de kits pide que se borren sus datos. Describe qué consulta de selección harías antes, qué borrarías y en qué orden, y qué impide hacerlo al revés.

Investiga

20. Localiza la documentación técnica oficial de un gestor relacional (LibreOffice Base u otro) o una guía de la Agencia Española de Protección de Datos sobre seguridad, copias o informes. Registra la fecha de consulta, el ámbito, la afirmación que sustenta y una limitación.
21. Lee en el BOE el artículo 32 del RGPD e indica qué medida cubre cada una de estas pruebas de tu base: copia restaurada, rol de lectura y registro de pruebas. Anota la fecha de consulta.

TU INVESTIGACIÓN AVANZA · PASO 11 · I11 · BASE RELACIONAL COMPROBADA

Con los datos ficticios o autorizados de tu investigación: 1) diagrama relacional con las tablas, qué representa una fila de cada una y sus claves; 2) catálogo de datos con archivo, tabla, campo, tipo, valores, clave y versión; 3) base con integridad referencial y registro de pruebas (alta válida, huérfana, fuera de rango, borrado con dependientes); 4) consultas de selección, agregación, JOIN con DISTINCT y una acción en copia con el antes y el después; 5) formulario probado; 6) roles, copia restaurada e informe para tu destinatario. Anota la decisión, el responsable y la fecha de revisión, y conserva el motivo de cada cambio de versión.

Comprueba tu paso: cada recuento dice qué cuenta, y cada prueba tiene su resultado esperado y el obtenido.

☐ Hecho ☐ Revisado ☐ Guardado

OJO PROFESIONAL

Documento: nota de comprobación de la base EK-2027/BD-v0.9

Documento ficticio de La Despensa de Nora, S.L., preparado por Unai el jueves 3 de junio de 2027 para que lo revise el equipo de análisis. Contiene **cinco errores**.

N.º	Apartado	Contenido
1	Estructura	Tres tablas con clave primaria; Encuestas enlaza con Personas por id_persona; Respuestas, con Encuestas por id_encuesta
2	Participación	El JOIN de las tres tablas devuelve 3 filas: han participado 3 personas
3	Integridad	La respuesta R06 con id_encuesta E07, que aún no existe, se ha aceptado; se enlazará cuando llegue la encuesta
4	Formulario	Se ha guardado un 8 en P1_interes (escala de 1 a 5); el formulario no dio ningún error
5	Prueba de acción	Para probarla, se ha cambiado el canal de E02 a teléfono directamente en la base de trabajo
6	Copia	Copia al cierre del día, 2027-06-03_estudio_kits_base_v0.9, en otro soporte; restaurada con los mismos recuentos
7	Exportación	Para Nora se exporta la tabla Respuestas completa con id_persona, zona y teléfono

Encargo. Señala los cinco errores, explica en cada uno qué regla de la unidad incumple y reescribe la línea para que la nota sirva. **Para pensar:** ¿qué prueba añadirías a la nota para que el error 3 no vuelva a pasar desapercibido?

Tabla de corrección

N.º	Por qué es un error	Corrección

AUTOEVALUACIÓN

1. En la tabla Respuestas del estudio, una fila representa...
a) una persona encuestada.
b) la respuesta a una pregunta de una encuesta.
c) un cuestionario completo.

2. Insertas R99 con id_encuesta E09, que no existe, y la base la acepta. Lo que falla es...
a) la clave primaria de Respuestas.
b) la regla de valores de 1 a 5.
c) la clave foránea, que no está definida o no está activada.

3. Antes de cambiar el canal de E02 con una consulta de acción debes...
a) copiar, ver qué filas cambian, ejecutar en la copia, comprobar y anotar.
b) ejecutar la acción y copiar después.
c) borrar E02 y darla de alta otra vez.

4. El formulario acepta un 7 en una escala de 1 a 5. Lo adecuado es...
a) avisar a quien teclea.
b) lista 1-5 en el formulario, regla CHECK en la tabla y prueba con un 7.
c) borrar a mano los valores imposibles al final.

5. Una copia de seguridad está comprobada cuando...
a) el archivo aparece en la carpeta de copias.
b) lleva la fecha en el nombre.
c) se ha restaurado aparte y da los mismos recuentos.

6. Personas P01 y P02; Encuestas E01 → P01 y E02 → P02; Respuestas R01 y R02 → E01, R03 → E02. El JOIN de las tres tablas da...
a) 3 filas, 2 encuestas y 2 personas.
b) 3 filas y 3 personas.
c) 2 filas y 2 personas.

7. En la base de Ruedalia añades E03 → P01 con R04 y R05. Ahora hay...
a) 5 respuestas, 3 encuestas y 3 personas.
b) 5 respuestas, 3 encuestas y 2 personas.
c) 3 respuestas, 3 encuestas y 2 personas.

8. «Tres filas tras unir tablas demuestran que hay tres entrevistados.» La corrección es...
a) es cierto si no hay filas duplicadas.
b) hay que contar las filas de la tabla Encuestas.
c) cada fila del JOIN es una respuesta; las personas se cuentan con COUNT(DISTINCT id_persona).

9. En la nota de Unai, la respuesta R06 → E07 aceptada se corrige...
a) activando la clave foránea, rechazando R06 y registrando la incidencia.
b) creando una E07 vacía para que encaje.
c) dejándola y avisando en el informe.

10. En tu entrega I11 decidiste no guardar teléfonos. Te haría revisarlo...
a) que la base pase de cien respuestas.
b) que el estudio prevea volver a contactar, con autorización, a las mismas personas.
c) que a Unai le resulte más cómoda una sola hoja.

Ya sé...

- ☐ distinguir fila, clave, consulta, formulario e informe.
- ☐ diseñar y comprobar la base del caso con sus claves y su integridad.
- ☐ transferir el procedimiento a la base de otro negocio, como Ruedalia.
- ☐ detectar el error de contar filas del JOIN como personas.
- ☐ entregar evidencia revisable: pruebas, recuentos, copia restaurada y versión.



UNIDAD 12

Informe, archivo y defensa

Bloque 4 · Análisis y comunicación de resultados

EN ESTA UNIDAD APRENDERÁS A…

- Ordenar problema, objetivo, método y resultados en un informe editable con resumen ejecutivo y anexo reproducible.
- Presentar tablas y gráficos con título informativo, unidad, base y fuente, sin exagerar diferencias.
- Separar hallazgo, interpretación y limitación, y poner en la conclusión los límites que cambian la decisión.
- Recomendar una acción proporcionada al nivel de detalle de quien decide y defenderla ante objeciones.
- Incorporar resultados, diccionario y versión a la base del SIM y comprobar que se pueden recuperar.

SITUACIÓN DE PARTIDA

Jueves 17 de junio de 2027. El equipo presenta en la trastienda los resultados del **Estudio de kits 2027**: 120 de las 200 personas de la muestra respondieron «sí, seguro» o «probablemente sí» a encargar el kit. Nora escucha de pie, con Olvido: «Mañana tengo que decidir».

- Unai ha preparado una diapositiva: «60 % del barrio quiere kits: 3.600 kits al mes».
- Olvido pregunta si decir «probablemente sí» es lo mismo que comprar.
- Adrián, que acaba mañana su formación, comprueba que la tabla y el gráfico salen de la versión archivada de los datos.

Nora te encarga el informe **EK-2027/IN-v1.0**: una recomendación para mañana, con lo que sostienen los datos y lo que aún no se sabe, y su paso al SIM.

PARA EMPEZAR

1. ¿Qué tiene que decidir Nora mañana y qué tiene que comprobar antes?
2. ¿Qué sabemos de verdad (120 respuestas favorables de 200) y qué estamos suponiendo en la diapositiva de Unai?
3. ¿Qué prueba permitiría revisar la cifra de 3.600 kits antes de enseñarla?
4. Si Nora solo puede leer una página, ¿qué tiene que estar en ella?



Objetivo y método

12.1



Resultados y gráficos

12.2



Conclusiones y límites

12.3



Recomendación

12.4



Transferencia al SIM

12.5

12.1 Objetivo y método

Un informe se escribe para que quien lo lee pueda evaluar la conclusión, no solo para que se la crea.

Para qué sirve un informe de investigación. El informe cierra el estudio. A sus cinco partes básicas (objetivo, metodología, resultados, conclusiones y limitaciones) se añaden tres piezas: un **resumen ejecutivo** para quien decide, un **anexo reproducible** para quien quiera comprobarlo y un **control de versiones** para saber de qué datos sale cada cifra.

El orden lógico. Problema → objetivo → método → resultados → conclusiones y límites → recomendación. Si el lector encuentra la recomendación sin el método, no sabe cuánto fiarse; si encuentra los resultados sin el objetivo, no sabe para qué sirven.

Tabla 1 · Estructura del informe EK-2027/IN-v1.0

Parte	Qué contiene en el estudio de kits	Para quién
Portada y versiones	Título, fecha, autoría, versión del informe y de los datos	Todos
Resumen ejecutivo	Decisión, hallazgo, intervalo, límite y recomendación	Nora
Problema y objetivo	¿Lanzar los kits? Medir la intención de encargo	Todos
Método	Población, selección, n = 200, campo, cuestionario, pregunta	Quien comprueba
Resultados	Tabla y gráfico de P1; favorables con su intervalo	Todos
Conclusiones y límites	Hallazgo, interpretación y límites materiales	Nora y equipo
Recomendación	Acción, condición y fecha de revisión	Nora
Anexos	Cuestionario, diccionario, consulta, datos, cálculos, cambios	Quien comprueba

El resumen ejecutivo. Una página que se entiende sin leer el resto: la decisión, el hallazgo con su base («120 de 200»), el intervalo, el límite que más pesa y la recomendación con su condición. No sustituye al método: lo resume y remite a él. Se escribe al final, con el informe cerrado, aunque vaya delante.

Ejemplo. Resumen ejecutivo de EK-2027/IN-v1.0 (borrador del equipo)

En mayo-junio de 2027 se preguntó a una muestra aleatoria simple de 200 hogares y oficinas del listado si encargarían el kit con entrega programada el próximo mes. Respondieron «sí, seguro» o «probablemente sí» 120 (60,0 %); el porcentaje en la población estaría entre el 53,21 % y el 66,79 % (95 % de confianza). Es intención declarada, no compra: solo el 24,0 % dijo «sí, seguro». Recomendación: piloto de cuatro semanas con oficinas y hogares H1 (cercanos, con prisa; unidad 1), revisado con las ventas reales.

El método y el anexo reproducible. El método dice lo necesario para juzgar la conclusión: población y marco, forma de selección, tamaño, fechas de campo, versión del cuestionario, pregunta literal y tratamiento de «no sabe». El **anexo reproducible** reúne los archivos con los que otra persona puede rehacer el 60 %: datos con su versión, diccionario EK-2027/DD-v1.0, consulta o fórmula y cálculo del intervalo. Si al rehacerlo no sale 120, el informe no está listo. Por eso el anexo no es un cajón de archivos: lleva un índice y una nota que dice en qué orden se abren y qué resultado debe salir en cada paso.

Herramientas. El informe se escribe en un procesador de textos con estilos de título, para que el índice se actualice solo, y las tablas y gráficos salen de la hoja de cálculo vinculada a los datos. Se entrega **editable** y, además, en PDF con la versión en el pie. En la portada, un **control de versiones**: «v0.9, 15/06, borrador para revisión; v1.0, 17/06, versión presentada». Si una cifra cambia después, se crea la v1.1 con el motivo, y la v1.0 se conserva.

ACTIVIDADES

1. Produce para el caso un informe editable y su resumen ejecutivo con método, resultados, límites y recomendación. Usa la estructura de la tabla 1 y comprueba que el resumen se entiende sin leer el resto.
2. Escribe el apartado de método del informe en un máximo de diez líneas. Señala qué dato del método necesitaría Nora para fiarse del 60 % y cuál solo necesita quien lo comprueba.

12.2 Resultados y gráficos

Una tabla o un gráfico sin base, unidad y fuente no se pueden comprobar.

La tabla de resultados. Una tabla estadística presenta los datos ordenados para leerlos y comprobarlos. Su **título informativo** dice el hallazgo, no solo el tema («Seis de cada diez encargarían el kit», mejor que «Pregunta 1»); indica la **unidad** (respuestas, %), la **base** (n), la **fuente** y la fecha, y las **notas materiales**: lo que cambia la lectura, como incluir «no sabe» en la base.

Tabla 2 · Seis de cada diez participantes declaran que encargarían el kit

(P1; base: 200)

Respuesta	Respuestas	%
Sí, seguro	48	24,0
Probablemente sí	72	36,0
Probablemente no	44	22,0
Seguro que no	28	14,0
No sabe / no contesta	8	4,0
Total	**200**	**100,0**

Fuente: encuesta del Estudio de kits, muestra aleatoria simple, campo de mayo-junio de 2027, datos v1.0 del 14/06/2027. Pregunta: «Si La Despensa ofreciera este kit con entrega programada, ¿lo encargaría en el próximo mes?». «No sabe / no contesta» se incluye en la base.

Elegir el gráfico. El gráfico sirve para ver de un golpe lo que la tabla dice con precisión. **Barras** para comparar categorías: en una escala, en su orden natural; en categorías sin orden, de mayor a menor. **Sectores**, solo para pocas partes de un total. **Líneas**, para una evolución en el tiempo. **Dispersión**, para dos variables numéricas. Los cuadros de detalle van al **anexo estadístico**.

Tabla 3 · Ficha del gráfico 1 del informe

Elemento	En el gráfico 1
Tipo	Barras horizontales, en el orden de la escala (de «sí, seguro» a «no sabe»)
Título informativo	«Seis de cada diez encargarían el kit; solo uno de cada cuatro, seguro»
Eje y unidad	Porcentaje de respuestas, desde 0
Base y fuente	n = 200; encuesta del Estudio de kits 2027, datos v1.0
Color	Favorables en un tono, desfavorables en otro, «no sabe» en gris

Tablas cruzadas.

Para comparar grupos, la tabla cruza dos variables y calcula el porcentaje **dentro de cada grupo**, con su propia base:

Tabla 4 · Favorables en P1 por tipo de participante

(campo de 2027)

Tipo	Favorables	Base	% de favorables
Hogares	84	130	64,6
Oficinas	36	70	51,4
Total	120	200	60,0

Los porcentajes se calculan sobre la base de cada fila, no sobre el total (84 es el 42,0 % de 200, pero el 64,6 % de los hogares). Se redondean a un decimal, y si una columna no suma exactamente 100,0 por el redondeo, se dice en una nota.

Cómo se exagera una diferencia. Con los datos de la tabla 4, los hogares dan un 64,6 % de favorables y las oficinas, un 51,4 %. Si el eje de las barras empieza en el 50 %, la barra de hogares mide 14,6 puntos y la de oficinas, 1,4: parece diez veces más larga, cuando el porcentaje es solo una cuarta parte mayor (64,6 ÷ 51,4 ≈ 1,26). Las barras empiezan en cero, sin tres dimensiones ni colores que destaquen lo que no es diferente, y cada grupo lleva su base.

CLAVE · BARRAS DESDE CERO Y BASE A LA VISTA

Un gráfico honesto deja ver la diferencia que hay, ni más ni menos: eje desde cero, unidad y base en el propio gráfico.

ACTIVIDADES

3. Con la hoja de cálculo, prepara la tabla y el gráfico de P1 con título informativo, unidad, base y fuente. Incluye la nota sobre «no sabe / no contesta».
4. Dibuja (o construye en la hoja) el gráfico de hogares y oficinas con el eje en el 50 % y después desde cero. Explica qué impresión da cada uno y escribe el título informativo correcto.

12.3 Conclusiones y límites

Hallazgo, interpretación y limitación son tres frases distintas, y el límite que cambia la decisión va en la conclusión.

Tres tipos de afirmación. Un **hallazgo** es lo que dicen los datos, con su base. Una **interpretación** es lo que el equipo infiere de él con un método y unos supuestos. Una **limitación** dice hasta dónde llega la conclusión. Mezclarlos es el error más frecuente de un informe: «el 60 % quiere el kit» convierte un hallazgo de respuestas en un deseo de toda la población.

Tabla 5 · Hallazgo, interpretación y limitación en el estudio de kits

Tipo	Afirmación del informe
Hallazgo	120 de 200 participantes (60,0 %) respondieron «sí, seguro» o «probablemente sí»
Interpretación	Con selección aleatoria simple y $n\cdot\hat{p}$ y $n\cdot\hat{q} \geq 10$, el porcentaje en la población está entre el 53,21 % y el 66,79 % (95 % de confianza, cobertura del procedimiento)
Limitación	Es intención declarada para el próximo mes, no compra; «probablemente sí» (36,0 %) pesa más que «sí, seguro» (24,0 %)

Los límites que cambian la decisión. Una limitación es **material** si, de ser cierta, cambiaría lo que Nora decide. Aquí lo son la distancia entre intención y compra, y que solo uno de cada cuatro participantes respondió con seguridad. Esos límites van en la conclusión y en el resumen ejecutivo; los que no cambian la decisión, como el redondeo, pueden ir en el anexo. Esconder en un anexo un límite material es tan engañoso como no escribirlo.

Las diferencias entre grupos, con su precisión. Cada grupo tiene menos casos que el total, así que su estimación es menos precisa: con 70 oficinas, el margen al 95 % sería de unos ±11,7 puntos (del 39,7 % al 63,1 %); con 130 hogares, de unos ±8,2 (del 56,4 % al 72,8 %). Los dos intervalos se solapan: el informe describe la diferencia, pero no recomienda dejar fuera a las oficinas por ella.

Escribir con prudencia. «Los datos indican», «en la muestra»; no «demuestra» ni «asegura». Cada conclusión cita su base.

Conclusiones relevantes, no ventas inventadas. La diapositiva de Unai multiplica el 60 % por los unos 6.000 hogares del barrio y lo llama «3.600 kits al mes». Tiene tres errores: aplica el resultado a una población distinta de la estudiada, convierte una intención en una compra y supone un encargo al mes por hogar. La conclusión relevante es otra: hay intención suficiente para **probar** el kit, no para dimensionar la producción.



Lo que aún no se sabe se escribe. El informe cierra la conclusión con una lista corta de incertidumbres y con la forma de reducirlas: cuántos de los que dicen «sí» encargan de verdad, con qué frecuencia y a qué hora de entrega. Una incertidumbre con su comprobación prevista es información para decidir, no una debilidad del informe.

ACTIVIDADES

5. Para el caso, escribe una conclusión con su hallazgo, su interpretación y su limitación material. Propón una conclusión alternativa basada solo en «sí, seguro» (48 de 200) y la condición que te haría preferirla. Usa la ficha de I12.

6. Clasifica como hallazgo, interpretación o limitación: a) «36 de 70 oficinas respondieron de forma favorable»; b) «Los hogares están más interesados que las oficinas»; c) «El campo fue en primavera, antes de las vacaciones»; d) «Entre el 53,21 % y el 66,79 % de la población declararía esa intención». Corrige la que esté mal formulada.

12.4 Recomendación y destinatario

Recomienda lo que los datos sostienen y al nivel de detalle que necesita quien decide.

El destinatario manda en el nivel de detalle. El mismo estudio se cuenta de tres maneras. Nora necesita una página con la recomendación, la condición y el riesgo; Olvido e Itxaso, qué cambia en el obrador y en la preparación de kits; quien revise el estudio, el método y el anexo. Cambia el detalle, no la conclusión ni sus límites.

Tabla 6 · Destinatarios del informe EK-2027/IN-v1.0

Destinatario	Qué necesita	Formato
Nora (decide)	Recomendación, condición, riesgo, revisión	Resumen y defensa oral
Olvido e Itxaso (operan)	Qué kits, para quién y cuántas semanas	Nota de una página
Revisión del estudio	Método, datos, consulta, cálculos, versión	Informe y anexos

Una recomendación proporcionada. Entre no hacer nada y lanzar los kits, los datos sostienen un paso intermedio y reversible: un **piloto de cuatro semanas**, del lunes 21 de junio al viernes 16 de julio de 2027, con oficinas y hogares H1 (cercanos con prisa), revisado con ventas reales el lunes 19 de julio. El criterio para ampliarlo se fija **antes** de empezar, para no interpretarlo a conveniencia: que a las dos semanas de su primer encargo repita al menos uno de cada dos clientes que prueben el kit (solo los que cumplieron el plazo).



La defensa. El jueves, Adrián enseña el gráfico impreso y el equipo responde a las preguntas. La defensa se prepara: se anticipan las objeciones sobre **selección**, **coste** y **precisión**, y se contesta con datos del informe, sin prometer lo que no se sabe. La exposición, de unos diez minutos, sigue el orden del informe y acaba con la recomendación y las preguntas.

Tabla 7 · Objeciones previstas y respuesta

Objeción	Respuesta preparada
Selección: «¿Solo habéis preguntado a los clientes?»	Muestra aleatoria simple de 200 del marco descrito en el método; quien no estaba en el marco no está representado, y se dice
Coste: «¿Y si el piloto sale caro?»	Cuatro semanas, con envases y raciones para lo encargado; si no se cumple el criterio, se para sin haber invertido en producción diaria
Precisión: «¿Es un 60 % exacto?»	Es una estimación: entre el 53,21 % y el 66,79 %; el extremo bajo no cambia la recomendación del piloto

Registrar la decisión. La defensa termina con un acta breve: qué se decide, con qué condición, quién se responsabiliza y cuándo se revisa. El viernes 18 de junio, Nora decide el piloto; Itxaso se encarga de registrar los encargos y la repetición; la revisión queda para el lunes 19 de julio. El acta se archiva con el informe.

Honestidad con quien encarga. El Código Internacional ICC/ESOMAR (edición de 2025), código de autorregulación de la profesión, pide transparencia sobre métodos, fuentes y limitaciones. Un informe que dice lo que el destinatario quiere oír, sin sus límites, no le ayuda a decidir. Consulta la versión vigente del código.

ACTIVIDADES

7. Redacta para Nora la recomendación del caso en cinco líneas como máximo, con su condición y su fecha de revisión. Añade una alternativa (lanzar ya o no hacer nada) y la condición de los datos que haría cambiar tu elección.
8. Olvido objeta: «Los que dicen "probablemente sí" no compran nunca». Prepara una respuesta con datos del informe y explica qué parte de la objeción comparte el equipo.

12.5 Actualización y transferencia

Un resultado archivado sin diccionario ni versión no se puede volver a consultar.

Del informe al SIM. El sistema de información de marketing (SIM) de La Despensa guarda los registros internos, la inteligencia del entorno y los resultados de investigación. Cerrar el estudio es **incorporar** a su base de datos los resultados agregados, el diccionario, la consulta y la versión, de forma que dentro de un año alguien pueda recuperar el 60 % y saber de dónde sale. Al SIM no van nombres ni contactos: van tablas agregadas. Lo que entra en el SIM lo consulta después cualquier persona del equipo, y debe entenderse sin quien lo hizo.

Un formato compatible. Los resultados se exportan en CSV UTF-8, legible en cualquier programa, con una fila por categoría de respuesta, y se importan en una tabla Resultados con clave foránea (id_estudio), como en la unidad anterior. La clave primaria de Resultados combina estudio, versión de los datos, pregunta y categoría: si alguien importa el mismo archivo dos veces, la base rechaza los duplicados.

Tabla 8 · Campos de la tabla Resultados del SIM

Campo	Ejemplo	Regla
id_estudio	EK-2027	clave foránea; parte de la clave
version_datos	v1.0	parte de la clave
pregunta	P1	código del diccionario; parte de la clave
categoria	si_seguro	código del diccionario; parte de la clave
respuestas	48	entero, 0 o más
base	200	igual en todas las filas de la pregunta

Verificar la consulta posterior. Tras importar, se comprueba que los datos se recuperan: SELECT SUM(respuestas) FROM Resultados WHERE id_estudio = 'EK-2027' AND version_datos = 'v1.0' AND pregunta = 'P1'; → 200, igual a la base. SELECT SUM(respuestas) FROM Resultados WHERE id_estudio = 'EK-2027' AND version_datos = 'v1.0' AND pregunta = 'P1' AND categoria IN ('si_seguro', 'probablemente_si'); → 120. Una segunda importación del mismo archivo se rechaza por clave repetida, y la suma sigue en 200.

La cadena de trazabilidad. Cada conclusión del informe tiene que poder seguirse hasta el dato que la sostiene:

Tabla 9 · Cadena conclusión ↔ gráfico ↔ consulta ↔ datos ↔ versión

Eslabón	En el estudio de kits	Comprobación
Conclusión	«60,0 % favorable (IC 95 %: 53,21-66,79 %)»	Coincide con el gráfico
Gráfico	Gráfico 1, P1, n = 200	Sale de la consulta guardada
Consulta	q_p1_favorables (120 de 200)	Se ejecuta sobre los datos archivados
Datos	200 respuestas depuradas, diccionario EK-2027/DD-v1.0	Archivo de análisis: una fila por persona
Versión	Datos v1.0 (14/06/2027) · informe IN-v1.0 (17/06/2027)	La misma en el informe y en el archivo

El archivo del estudio. Una carpeta por estudio, con fecha y versión en cada nombre: plan, cuestionario, datos brutos sin tocar, datos depurados, diccionario, consultas, informe y registro de cambios. Los contactos se borran cuando ya no hacen falta; los resultados agregados y el informe se conservan para compararlos luego.

Presentar desde la base. Una consulta guardada en la base del SIM devuelve el cuadro de P1 con su base y su versión cuando se pide, sin copiar cifras a mano del PDF.

Actualización. Cuando lleguen las ventas del piloto, se añaden como datos nuevos, con su versión, a los registros internos del SIM. Nunca se sobrescribe el resultado anterior: si se corrige un dato, se crea una versión nueva y se anota el motivo en el registro de cambios. Así se sabe qué cifra vio Nora al decidir y qué se corrigió después y por qué motivo.

LA NORMA DICE

RGPD, art. 5.1.e): los datos personales se mantendrán de forma que permitan identificar a los interesados «durante no más tiempo del necesario» («limitación del plazo de conservación»). Considerando 162: el fin estadístico implica que el resultado no sean datos personales, «sino datos agregados». Consulta la versión vigente.

EN TU SECTOR

Marketing y publicidad: el informe de una campaña archiva con cada resultado su medio, su periodo y su base, para compararla con la siguiente. **Gestión de ventas y espacios comerciales:** el resultado de una prueba de surtido entra en el sistema con la tienda, las semanas y la versión de las ventas.

ACTIVIDADES

9. Importa a una base del SIM los resultados de P1, el diccionario y la versión. Verifica con dos consultas posteriores que la suma es 200 y que los favorables son 120, e intenta importar el archivo otra vez.
10. Tras una importación, la suma de P1 da 400. Explica qué ha pasado, cómo lo habría impedido la clave primaria y qué anotarías en el registro de cambios.

CASO RESUELTO · PASO A PASO

El informe del Estudio de kits

Caso ficticio. El jueves 17 de junio de 2027 el equipo de análisis entrega a Nora el informe EK-2027/IN-v1.0 con el resultado de la encuesta: 120 favorables de 200.

Paso 1. Decisión, datos y alcance. Nora decide el viernes 18 si lanza los kits con entrega programada. Datos: P1 de una muestra aleatoria simple de 200 hogares y oficinas, campo de mayo-junio, datos v1.0 del 14/06: 48 «sí, seguro», 72 «probablemente sí», 44 «probablemente no», 28 «seguro que no» y 8 «no sabe». Alcance: intención declarada para el próximo mes, no ventas.

Paso 2. Procedimiento. 1) Tabla de P1 con unidad, base 200 y fuente; los porcentajes, sobre 200. 2) Favorables: 48 + 72 = 120 → $\hat{p}$ = 120 ÷ 200 = 0,60 (60,0 %). 3) Condiciones: $n\cdot\hat{p}$ = 120 ≥ 10 y $n\cdot\hat{q}$ = 80 ≥ 10. 4) Intervalo al 95 %: EE = √(0,60 · 0,40 ÷ 200) = 0,034641; E = 1,96 · 0,034641 = 0,0679; 60,0 % ± 6,79 puntos → 53,21 %-66,79 %. 5) Gráfico de barras desde cero con base y fuente. 6) Resumen ejecutivo, anexo de método y recomendación condicionada. 7) Resultados, diccionario, consulta y versión al SIM.

Paso 3. Comprobación e interpretación. La tabla suma 200 y 100,0 %; la consulta del SIM devuelve 120 de 200, y el informe y el archivo citan la misma versión de los datos. Permite concluir que, con este procedimiento, la intención de encargo está entre el 53,21 % y el 66,79 %. **No** permite concluir cuántos kits se venderán: intención no es compra y solo el 24,0 % respondió «sí, seguro». Por eso no se calculan ventas previstas multiplicando el 60 %.

RESULTADO

60,0 % favorable (120 de 200), con un intervalo del 53,21 % al 66,79 % al 95 % y sus supuestos. Recomendación condicionada: piloto de cuatro semanas con oficinas y hogares H1, del 21/06 al 16/07/2027, con un criterio fijado antes y revisión con las ventas reales el 19/07. Conclusión, gráfico, consulta, datos y versión, enlazados en el SIM.

Tabla 10 · Control del informe antes de entregarlo

Qué se comprueba	Resultado
Suma de la tabla de P1	48 + 72 + 44 + 28 + 8 = 200 (100,0 %)
Favorables y condiciones	120 de 200 = 60,0 %; 120 y 80 ≥ 10
Intervalo al 95 %	60,0 % ± 6,79 → 53,21 %-66,79 %
Trazabilidad	Consulta del SIM = 120; versión v1.0 en informe y archivo

Comprueba que lo has entendido. a) Si el gráfico dijera n = 192 porque se quitó «no sabe», ¿qué cambiaría en la tabla y en el texto? b) ¿Por qué el 53,21 % no impide recomendar el piloto? c) ¿Qué dato del piloto haría revisar la recomendación?

PRÁCTICA

El informe de Ruedalia

Caso ficticio. Txema quiere saber si a las empresas del polígono les interesa una revisión mensual de sus bicicletas en la propia empresa. Del lunes 14 al viernes 25 de junio de 2027 dejó una pregunta en las facturas y en un cartel del taller: «¿Contrataría la revisión mensual en su empresa?». Contestaron **20** clientes de empresa, por su cuenta, y **14** dijeron que sí. Txema quiere poner en el folleto «El 70 % de las empresas del polígono quiere nuestro servicio» y que Ousmane reserve agenda para 14 contratos.

Encargo. Con el procedimiento del caso resuelto: 1) Tabla de resultados con unidad, base, fuente y periodo, y un gráfico adecuado con título informativo. 2) Porcentaje y conclusión con su alcance: ¿se puede calcular un intervalo como en el caso? Justifícalo con la forma de selección y con las condiciones $n \cdot \hat{p} \geq 10$ y $n \cdot \hat{q} \geq 10$. 3) Resumen ejecutivo para Txema con hallazgo, límite y una recomendación proporcionada. 4) Registro de resultados para el SIM de Ruedalia con estudio, versión, pregunta, categoría, respuestas y base.

Plantilla IC-P12 · Informe de Ruedalia

Campo	Tu respuesta
Datos y fuentes con unidad y periodo	
Procedimiento o cálculo	
Resultado y evidencia	
Decisión y limitaciones	

Para pensar. ¿Qué le dirías a Txema sobre la frase del folleto y sobre la agenda de Ousmane? ¿Qué tendría que hacer para poder hablar de «las empresas del polígono»?

Lista de cotejo

☐ Tabla y gráfico con base 20, fuente y periodo.
☐ Porcentaje dicho como «de quienes respondieron».
☐ Sin margen de error, con el motivo.
☐ No confunde 14 respuestas con 14 contratos.
☐ Recomendación proporcionada y registro para el SIM con versión.

LECTURA Y DEBATE

La víspera



Relato ficticio. A las nueve de la noche del jueves, con la persiana medio bajada, Nora extendió el informe sobre el mostrador. En la primera página estaba todo lo que quería oír: seis de cada diez. Sacó la libreta y empezó a echar cuentas. Si seis de cada diez clientes encargaban un kit, había que pedir ya cajas isotérmicas para tres meses, reservar dos mañanas del obrador y anunciar los kits en el escaparate el lunes.

Antes de cerrar la libreta, pasó la página del resumen y fue al método, como le había pedido el equipo. Allí estaba la tabla completa: 48 personas habían dicho «sí, seguro»; 72, «probablemente sí». En la conclusión, una frase subrayada: «Es intención declarada, no compra». Recordó a cuántos clientes les había oído decir «un día de estos lo pruebo» y cuántos lo habían probado.

Tachó «tres meses» y escribió «cuatro semanas». Las cajas, para lo que se encargara; el escaparate, sin anunciar nada hasta ver si quien probaba el kit repetía. Al día siguiente no dijo «lanzamos»; dijo «probamos, y en julio miramos las ventas».

Preguntas. 1) ¿Qué afirmación del caso necesitaba una comprobación antes de convertirse en decisión? 2) ¿Qué información pedirías primero y por qué cambiaría la respuesta? 3) Propón una actuación provisional distinta de la de Nora que permita aprender sin ocultar la incertidumbre.

APRENDE CON LA IA

Unai pide a un asistente de IA que le ayude a acortar el informe para Nora. Esta es la respuesta:

RESPUESTA DE LA IA

«Un informe convincente puede omitir la metodología para no confundir al cliente. Quédate con el 60 % en grande y la recomendación: cuanto más limpio, más fácil de aprobar.»

Tu tarea. Localiza el error, contrástalo con el procedimiento de la unidad (resumen ejecutivo, anexo reproducible y límites materiales) y escribe una respuesta corregida que diga también su límite. Conserva la prueba de tu corrección: qué parte del método cambia la lectura del 60 %.

Ojo con la IA. Un asistente puede resumir un texto, pero no sabe qué límite es material para tu decisión: eso lo decide el equipo con los datos.

ACTIVIDADES FINALES

Comprueba

11. Resume una distinción decisiva de esta unidad (por ejemplo, hallazgo frente a interpretación, o resumen ejecutivo frente a anexo) y aplícala al caso resuelto.
12. ¿Qué elementos debe llevar un gráfico para que se pueda comprobar? ¿Por qué un eje de barras que no empieza en cero exagera una diferencia?

Aplica

13. Repite la práctica de Ruedalia cambiando una condición relevante: Txema recibe 10 respuestas voluntarias más, de las que 5 son favorables. Señala qué resultados cambian (porcentaje, base, $n \cdot \hat{q}$) y cuáles se mantienen (alcance de la conclusión, margen de error).
14. Reescribe el resumen ejecutivo del caso para Olvido e Itxaso: qué cambia en el obrador y en la preparación de kits durante el piloto, sin perder la condición ni los límites.

Investiga

15. Localiza una fuente primaria o documentación técnica sobre actualización y transferencia de datos (documentación oficial de un gestor de bases de datos o de una hoja de cálculo sobre importar CSV, o una guía de la Agencia Española de Protección de Datos sobre conservación). Registra la fecha de consulta, el ámbito, la afirmación que sustenta y una limitación.
16. Busca en la web de la Cámara de Comercio Internacional (ICC) la edición vigente del Código Internacional ICC/ESOMAR y anota qué dice, con tus palabras, sobre la transparencia de métodos y limitaciones. Indica la fecha de consulta.

TU INVESTIGACIÓN AVANZA · PASO 12 · I12 · INFORME Y TRANSFERENCIA AL SIM

Con los datos ficticios o autorizados de tu investigación: 1) informe editable con resumen ejecutivo, método, resultados, conclusiones con sus límites materiales y recomendación; 2) tabla y gráfico con título informativo, unidad, base y fuente; 3) anexo reproducible (datos con versión, diccionario, consulta y cálculos); 4) resultados, diccionario y versión importados a la base del SIM, con la consulta posterior verificada; 5) cadena conclusión ↔ gráfico ↔ consulta ↔ datos ↔ versión. Anota la decisión, el responsable y la fecha de revisión. No basta una presentación sin archivos de cálculo, registros o pruebas.

Comprueba tu paso: otra persona rehace tu resultado principal con tus anexos y le sale lo mismo.

☐ Hecho ☐ Revisado ☐ Guardado

OJO PROFESIONAL

Documento: ficha de resultados de la encuesta en redes

Documento ficticio de La Despensa de Nora, S.L., preparado por Unai el martes 15 de junio de 2027 para archivarlo en el SIM. Contiene **tres errores**.

N.º	Apartado	Contenido
1	Estudio	Encuesta en las redes sociales de la tienda, del 1 al 11 de junio; respuesta voluntaria
2	Pregunta	«¿Encargaría un kit con entrega programada?» (sí / no)
3	Resultado	36 de 48 respuestas (v1.0) fueron «sí»: el 75 % del barrio quiere los kits
4	Gráfico	Barras «sí» y «no» en porcentaje, con el eje desde cero, sin base ni fuente
5	Datos	Archivo `2027-06-11_redes_kits_v1.0`, con diccionario
6	Informe	Las cifras del informe salen de `2027-06-14_redes_kits_v1.1` (se quitaron dos respuestas repetidas)
7	Uso	Resultado agregado; sin nombres ni usuarios de las redes

Encargo. Señala los tres errores, explica qué regla incumple cada uno y reescribe la ficha para que se pueda archivar. **Para pensar:** ¿qué comprobación añadirías al cerrar cada informe para que el error de versión no se repita?

Tabla de corrección

N.º	Por qué es un error	Corrección

AUTOEVALUACIÓN

1. Nora solo leerá una página. El informe debe...

a) quedarse en una página y suprimir el método.
b) llevar un resumen ejecutivo con hallazgo, límite y recomendación, y el método y los anexos detrás.
c) poner los límites solo en el anexo para no confundir.

2. Un gráfico de barras compara 64,6 % y 51,4 % con el eje desde el 50 %. Lo correcto es...

a) mantenerlo, porque las cifras están escritas.
b) usar sectores en tres dimensiones.
c) empezar el eje en cero e indicar la base de cada grupo.

3. «El 60 % del barrio comprará el kit» mezcla...

a) un hallazgo de respuestas con una compra de toda una población.
b) dos limitaciones.
c) un intervalo con una media.

4. Con 120 de 200 favorables y solo un 24,0 % de «sí, seguro», la recomendación proporcionada es...

a) lanzar con producción diaria.
b) un piloto con criterio fijado antes y revisión con ventas reales.
c) no hacer nada hasta tener una muestra de 1.000.

5. Tras importar dos veces el mismo CSV, la suma de P1 en el SIM da 400. Lo que falla es...

a) la clave primaria de Resultados, que debía impedir duplicados.
b) el diccionario.
c) la base 200, que hay que cambiar a 400.

6. En el caso, el intervalo de favorables al 95 % es...

a) 60,0 % ± 3,46 puntos.
b) 53,21 %-66,79 %.
c) 50,0 %-70,0 %.

7. Ruedalia: 14 de 20 respuestas voluntarias dicen «sí». Lo correcto es...

a) 70 % de las empresas del polígono, ± 20 puntos.
b) 70 % de quienes respondieron, sin margen de error.
c) 14 contratos asegurados.

8. «Un informe convincente puede omitir la metodología.» La corrección es...

a) cierto si el cliente no la pide.
b) la metodología va en el resumen ejecutivo completa.
c) el resumen puede ser breve, pero el método y los límites materiales tienen que estar y poder comprobarse.

9. En la ficha de Unai, el informe usa la v1.1 y el archivo, la v1.0. Se corrige...

a) archivando la v1.1 con su motivo de cambio y citándola en el informe.
b) borrando la v1.0.
c) dejando la v1.0, porque es la primera.

10. En tu entrega I12 recomendaste un piloto. Te haría revisarlo...

a) que el informe tenga más páginas.
b) que en las primeras semanas casi nadie repita el encargo.
c) que a Unai le guste más lanzar ya.

Ya sé...

- ☐ distinguir informe, gráfico, conclusión, recomendación y transferencia.
- ☐ elaborar el informe del caso: tabla, gráfico, intervalo y recomendación.
- ☐ transferir el procedimiento al informe de otro negocio, como Ruedalia.
- ☐ detectar una conclusión no sustentada, como «el 60 % del barrio comprará».
- ☐ entregar una evidencia revisable: anexos, consulta del SIM y versión.

josesande.com