

Métodos cuantitativos I:

Clase 3

Sebastián Muñoz Tapia
Mail: semunoz@uahurtado.cl

Lo que veremos

01

Sintesis

02

Muestreo

03

Ejercicio en clase

04

Ejercicio en casa

01

Síntesis de una investigación por encuesta



02

Muestreo

2.0 Formas de seleccionar parar “representar” totalidad



Encuesta CEP
DESDE 1987



ABRIL-MAYO 2022

Estudio Nacional de Opinión Pública

Encuesta CEP 86

www.cepchile.cl/encuestaCEP



Datos de la muestra

UNIVERSO QUE SE INCLUYE

Población de 18 años y más (urbana y rural) residente a lo largo de todo el país (a excepción de Isla de Pascua, Juan Fernández y la Antártica).

COBERTURA DEL UNIVERSO

Toda la población de 18 años y más en base al Censo de 2017 (a excepción de Isla de Pascua, Juan Fernández y la Antártica).

MUESTRA

1.355 personas fueron entrevistadas en sus hogares, en **126 comunas** del país.

El método de muestreo fue estratificado (por región y zona urbana/rural), aleatorio y probabilístico en cada una de sus tres etapas (manzana-hogar-entrevistado). No se utilizaron reemplazos.

El nivel de respuesta fue de **61,1%** lo que corresponde a las 1.355 entrevistas completas sobre la muestra original de 2.216 entrevistas.

NIVEL DE PRECISIÓN

El error muestral se estima en **± 2,9%** considerando varianza máxima, un nivel de confianza **95%** y un efecto del diseño estimado de 1,2.

INSTRUMENTO

Entrevistas cara a cara, basadas en un cuestionario estructurado.

FECHA DE TERRENO

La recolección de datos se efectuó entre el **13 de abril** y el **29 de mayo del 2022**.

Mayores de 18:
13.562.821
(CENSO 2017)

1355 de
14.000.000

=0,01%

¿Cómo es que con tan pocos casos se pueda intentar decir lo que pasa en todo Chile?

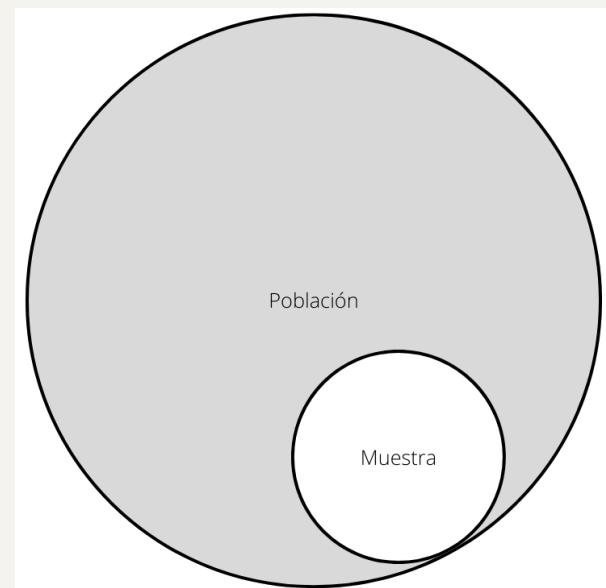
2.0 Formas de seleccionar para “representar” totalidad

- Al preparar una sopa, uno saca una pequeña cucharada para ver el nivel general de sal y aliños. ¿Por qué una pequeña cucharada puede “representar” la totalidad?
- ¿Pero si la cucharada se saca de un lugar que no está bien revuelto?
- ¿Cómo seleccionar para que la pequeña parte (la “muestra”) represente a la totalidad (la “población”)?



2.1 ¿Qué es el Muestreo?

- es el procedimiento mediante el cual se selecciona una **muestra** de una **población**, con el fin de conocer características de esa población.
- **Población:** conjunto de elementos del cual queremos extraer información, ubicado en un tiempo y espacio determinado.
- **Muestra:** unidades de observación concretas desde las cuales extraeremos la información para conocer a la población origen de la muestra



2.2 ¿Por qué estudiamos muestras y no poblaciones?

- **Recursos:** estudiar una muestra es más rápido y económico
 - **Precisión:** Una muestra **puede** ser más precisa en la estimación de los valores poblacionales que un CENSO.
-
- ¿Cómo se explica esta paradoja?
 - Errores ajenos al muestreo:
 - Errores del muestreo:

2.3 Conceptos claves de muestreo

Inferencia

Estimación

Parámetro y
Estadístico

Representatividad

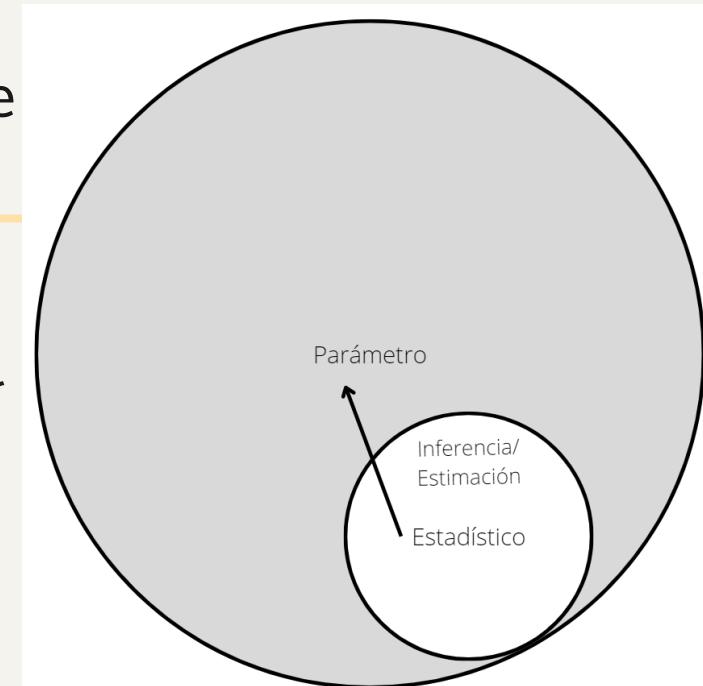
Azar

Errores de
muestreo

Muestreo
probabilístico y
no probabilístico

- Hay dos tipos de muestreos: **probabilísticos y no probabilísticos**
- El fin de un muestreo probabilístico **es utilizar información de una muestra representativa** para explorar, describir o explicar las propiedades de la población origen de la muestra.
- ¿Cómo logramos esto?....

- **Estadística inferencial:** conjunto de procedimientos estadísticos que permiten generalizar los resultados de la muestra a la población origen de la muestra.
- **Estimación:** procedimiento de inferencia que utiliza datos muestrales (estadísticos) para estimar un parámetro poblacional.
- **Parámetro poblacional:** valor desconocido de la población que es estimado por un valor conocido de la muestra.
- **Estadístico:** valor calculado a partir de los valores muestrales con el propósito de estimar el valor de un parámetro en la población.



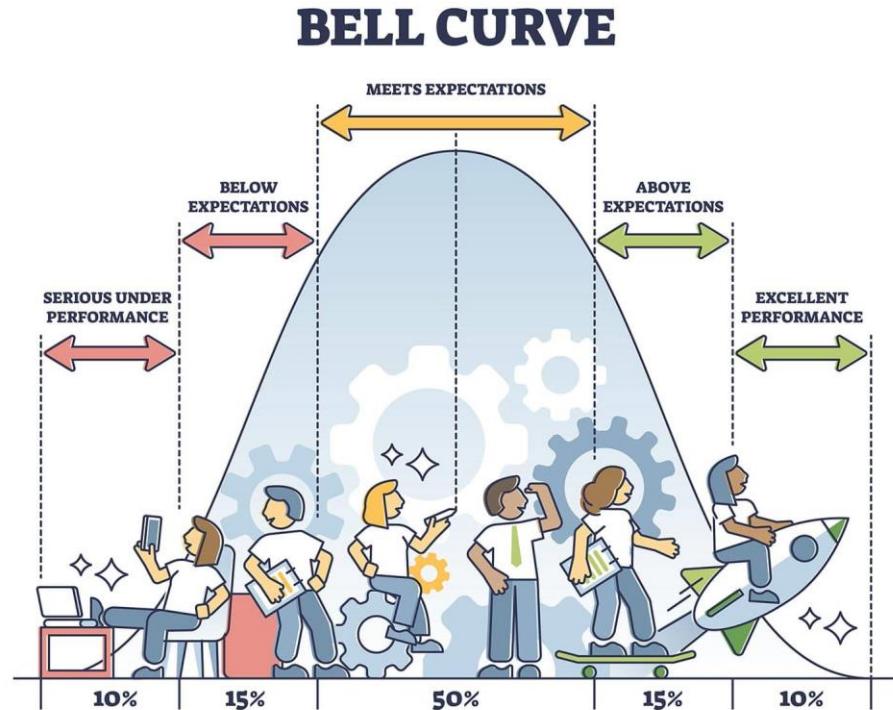
- Muestra:

- “no es cualquier parte de la población”
- Metafóricamente: “es una maqueta de la población en las que están reproducidos sus atributos a una escala menor”.

- **Representatividad:** la cualidad de reproducir a escala los atributos de la población, y en consecuencia, permitir inferir a la población origen de la muestra
- Sabemos que la representatividad es posible, gracias a **dos leyes** estadísticas esenciales:
 - **El teorema del límite central** → La distribución de medias muestrales extraídas de forma **aleatoria** de una población, se aproxima a la **distribución normal** a medida que **aumenta el tamaño de la muestra**.
 - **La ley de los grandes números** → la diferencia entre el estadístico muestral y el parámetro poblacional tiende a 0 cuando el **tamaño de la muestra tiende al infinito**

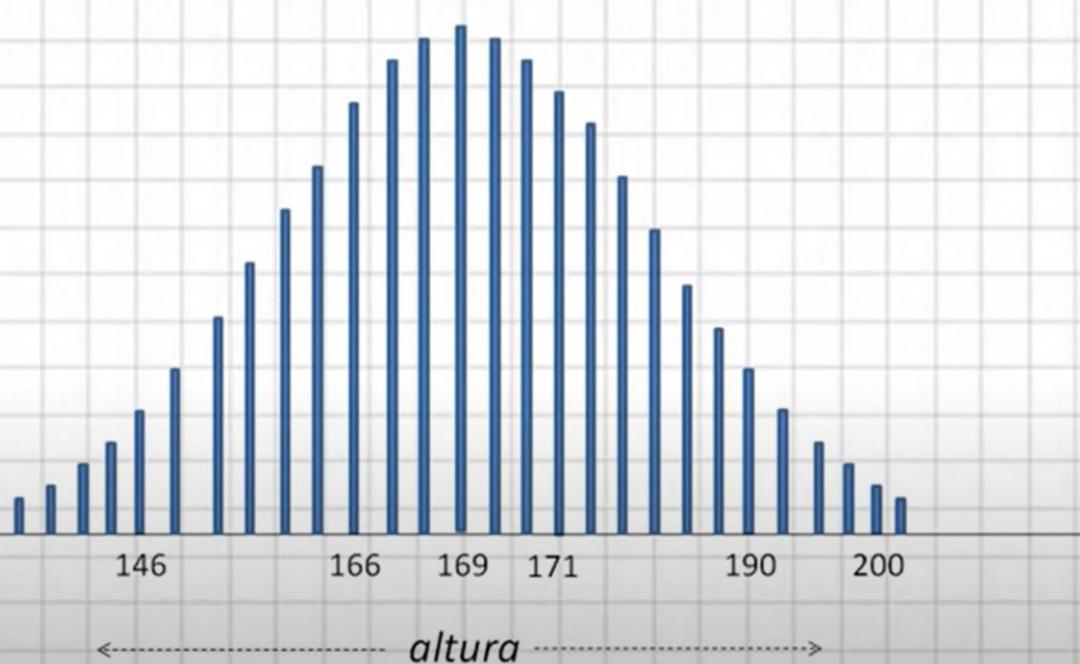
Distribución Normal

- Saco aleatoriamente distintas muestras, gran cantidad de esas muestras se acercarán a la curva normal.
- Ej.
 - Media de Altura muestra 1;
 - Media altura muestra 2;
 - Media muestra...
 - Media muestra N
- Se organizarán en una curva normal



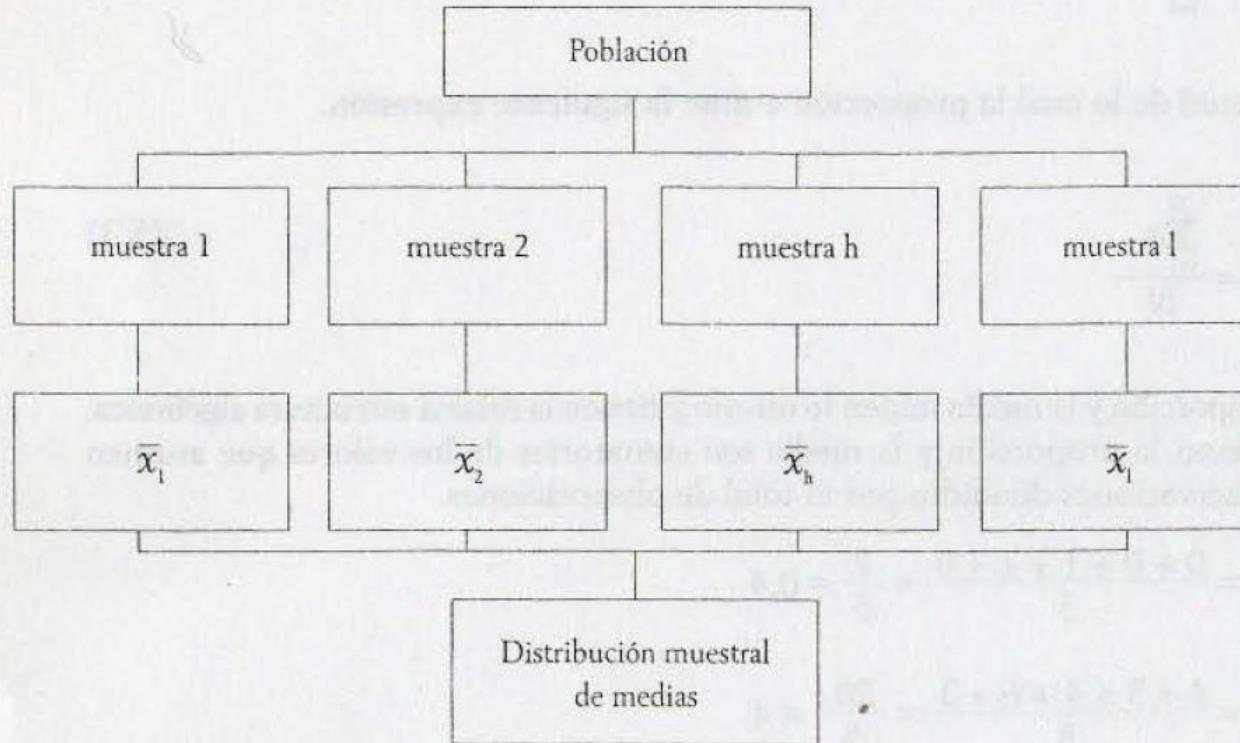
QUÉ ES LA DISTRIBUCIÓN NORMAL

nº de personas



- Distribución teórica de variables continuas
- Es una distribución simétrica (moda, mediana y media coinciden)
- Permite comparar otras distribuciones respecto a la forma de ésta
- Casi todo el mundo está en el medio

Gráfico 2. Distribución muestral de medias.

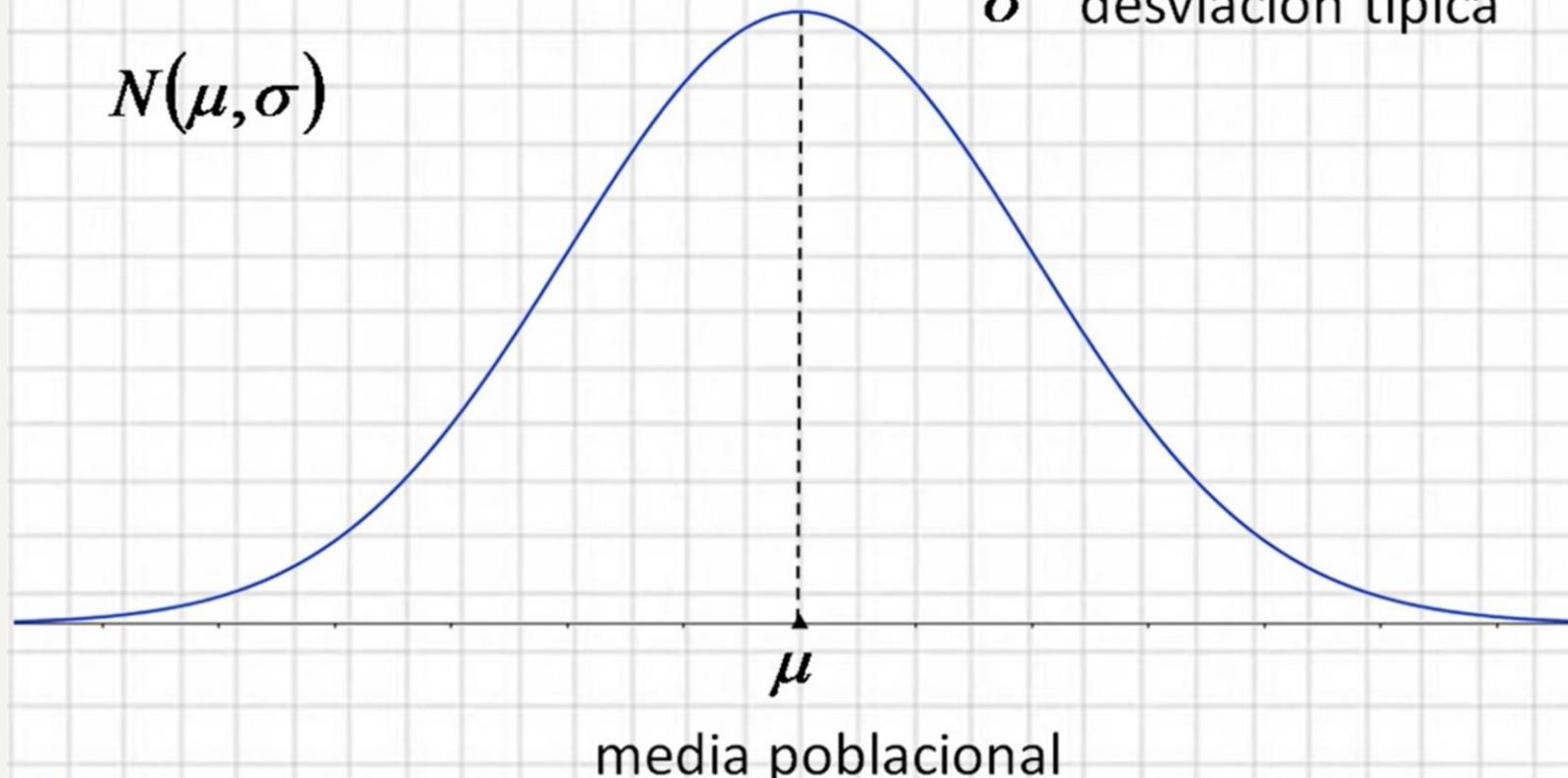


QUÉ ES LA DISTRIBUCIÓN NORMAL

Distribución Normal

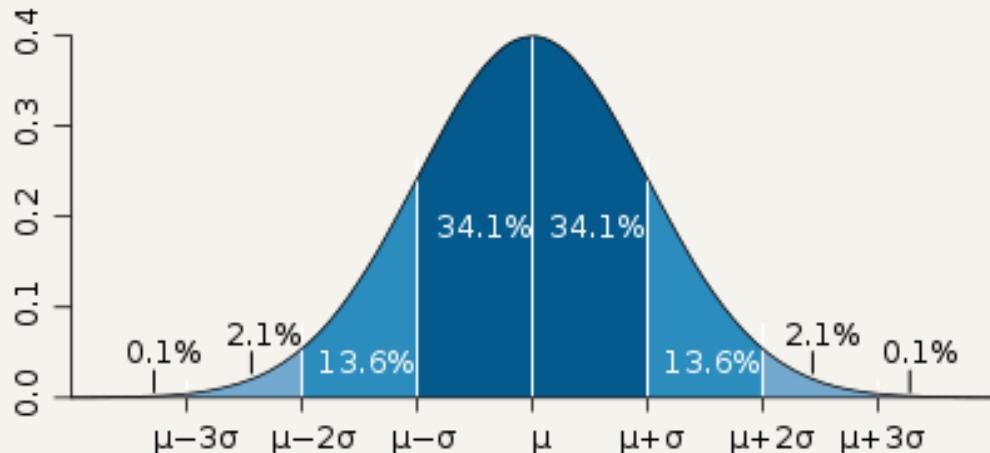
$$N(\mu, \sigma)$$

σ desviación típica



Supuesto Normalidad

- La **estadística paramétrica** involucra supuestos de normalidad en la distribución muestral de los valores de interés (ej: la media).
- Esto es importante para calcular áreas bajo la curva → significación estadística (probabilidad de que la relación que se encuentra en la muestra se encuentra en la población)
- El supuesto de normalidad suele cumplirse cuando $n > 30$ casos, aún si la distribución de la variable de interés no es normal en la muestra o la población



Datos de la muestra

UNIVERSO QUE SE INCLUYE

Población de 18 años y más (urbana y rural) residente a lo largo de todo el país (a excepción de Isla de Pascua, Juan Fernández y la Antártica).

COBERTURA DEL UNIVERSO

Toda la población de 18 años y más en base al Censo de 2017 (a excepción de Isla de Pascua, Juan Fernández y la Antártica).

MUESTRA

1.355 personas fueron entrevistadas en sus hogares, en **126 comunas** del país.

El método de muestreo fue estratificado (por región y zona urbana/rural), aleatorio y probabilístico en cada una de sus tres etapas (manzana-hogar-entrevistado). No se utilizaron reemplazos.

El nivel de respuesta fue de **61,1%** lo que corresponde a las 1.355 entrevistas completas sobre la muestra original de 2.216 entrevistas.

NIVEL DE PRECISIÓN

El error muestral se estima en **± 2,9%** considerando varianza máxima, un nivel de confianza **95%** y un efecto del diseño estimado de 1,2.

INSTRUMENTO

Entrevistas cara a cara, basadas en un cuestionario estructurado.

FECHA DE TERRENO

La recolección de datos se efectuó entre el **13 de abril** y el **29 de mayo del 2022**.

Mayores de 18: 13.562.821
(CENSO 2017)

1355 de
14.000.000
=0,01%

¿Cómo es que con tan pocos casos se pueda intentar decir lo que pasa en todo Chile?

- a) Aleatorio, probabilístico
- b) Error (conocido)



“Sin embargo, el azar no es garantía de representatividad. Siempre hay errores asociados a la estimación.”

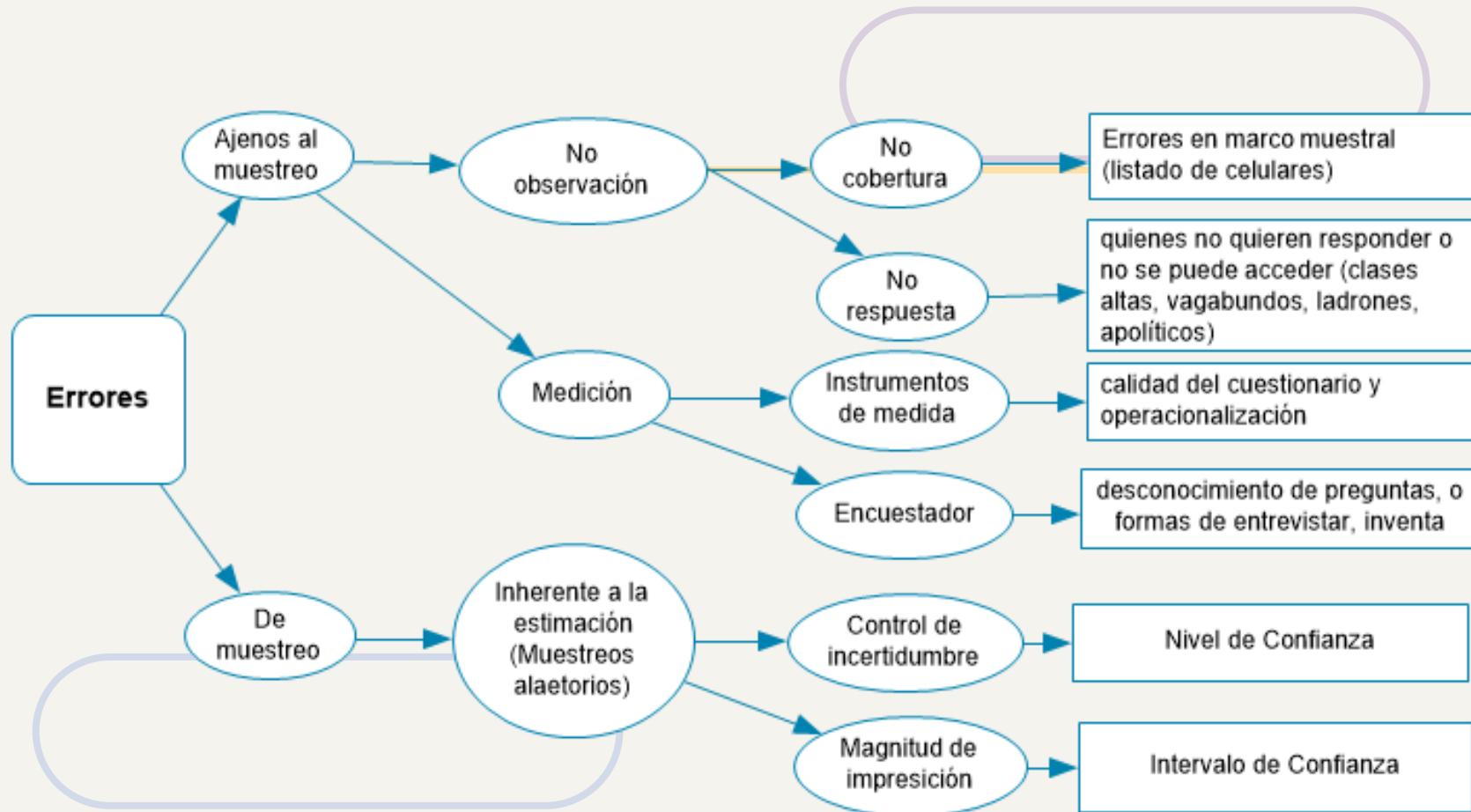
—¿Qué errores se imaginan más allá del estadístico?



Tipos de Error

- a. **Error de muestreo** → se producen por la **naturaleza de la inferencia**. Es la diferencia entre el parámetro poblacional y el del estadístico muestral, al extraer una muestra aleatoria.
- La magnitud del error de muestreo
 - Disminuye cuando la muestra es grande
 - Aumenta cuando la **dispersión** de las variables sometidas a medición es grande (cuando la población es muy heterogénea).
- ¿Los censos tienen errores de muestreo?

- **b. Errores ajenos al muestreo (errores sistemáticos)** → se originan en el conjunto de operaciones que se realizan en el proceso de recogida y almacenamiento de información
 - **Errores por no observación:** observación incompleta de las unidades que componen la población objetivo.
 - **Errores de medición:** pueden haber ‘errores de operacionalización’ y puede haber ‘efecto entrevistador’
- Los censos tienen más errores ajenos al muestreo que una muestra....



Ahora podemos entender la paradoja... ¿por qué una muestra puede ser más precisa que un censo?

- La experiencia enseña que los errores de muestreo son menores a los errores ajenos al muestreo.
- Se ha probado que una muestra bien diseñada permite estimar el parámetro poblacional con un error asociado menor que el producido en un censo.

2.4 Cómo se diseña una muestra

- Primer paso: definición del **marco muestral**
 - El ‘marco muestral’ son los insumos utilizados para identificar cada una de las unidades de muestreo.
 - Permite enumerar las unidades de muestreo para su posterior selección.
 - Suelen ser: listados de individuos (clientes, etc.); listados de manzanas, distritos, mapas y planos.
- En lo posible, se busca que el marco:
 - Sea completo
 - Esté actualizado
 - No debe incluir unidades que no correspondan a la población que se analiza
 - Debe contener información suplementaria que ayude a localizar a las unidades seleccionadas
 - Debe ser fácil de utilizar

Listado de
estudiantes de
Antropología
Universidad
Alberto Hurtado:
209 personas

Nº	Apellido Paterno	Apellido Materno	Nombres	Año Admisión				
2	4 TAPIA	VALLADARES	TAMARA JACQUELINE	2011	145	79 AGUILAR	TRONCOSO	SOFÍA VALENTINA
3	6 HERNANDEZ	HERNANDEZ	ALEJANDRO SALVADOR	2012	146	30 ALDANA	KURTE	CRISTOBAL FRANCISCO
4	11 COO	SANHUEZA	SOFIA INÉS ADRIANA	2014	147	7 ALVAREZ	REYES	MATIAS FERNANDO
5	1 OLIVA	CORNEJO	VALENTINA SOLEDAD	2014	148	129 ASTUDILLO	RAMÍREZ	NAYARETH ROCÍO IGNACIA
6	18 URRIOLA	ABALLAI	CATALINA ANDREA	2014	149	204 BUSTOS	PALMA	JAVIERA VIVIANA
7	23 PEÑA	SOTO	JAVIERA CONSTANZA	2015	150	12 CONCHA	CAICEDO	GABRIEL ROBERTO
8	29 CAQUISANE	TAPIA	JAVIERA FERNANDA	2016	151	144 CORNEJO	ROJAS	PABLO TOMÁS ARIEL
9	39 LAGOS	CARTES	CAROLINA ANDREA	2016	152	185 ESCAFI	BASULTO	SAMANTHA MICHELLE
10	9 LAS HERAS	TAPIA	IVAN DAVID	2016	153	104 ESPINOZA	MORALES	MARÍA JOSÉ
11	34 OPORTO	MORAN	NAYARET CATALINA	2016	154	50 FIGUEROA	ZAMORANO	CATALINA JAVIERA
12	48 ORTIZ	LATAPIAT	MARTINA ANGELICA	2016	155	219 FOGARTY	CHRYSTAL	SHEILA MAUREEN
13	26 PEREZ	JIMENEZ	MARIA IGNACIA	2016	156	181 FREDES	MORA	FERNANDA IGNACIA
14	16 VERA	REIMÁN	CLAUDIA ANTONIA	2016	157	94 GONZÁLEZ	NAVARRETE	ESTEFANO FABIÁN
15	40 ARANZAES	HERRÁN	CAROLINA ANTONIA	2017	158	175 HENRÍQUEZ	BRIONES	PAULINA IGNACIA
16	20 BASTÍAS	ÁLVAREZ	FELIPE ANDRÉS	2017	159	195 HERRERA	CATALAN	JAVIERA VIOLETA
17	15 CAUCHUPAN	GAJARDO	JAVIERA VALENTINA PAZ	2017	160	184 MATTE	DÍAZ	ELISA
18	10 COX	VERDUGO	MATIAS IGNACIO	2017	161	126 MAYORA	PICKER	NICOLAS
19	37 FIERRO	ESCOBAR	ANDRES SIMON	2017	162	132 MELÉNDEZ	HENRÍQUEZ	JOAQUÍN ANDRÉS
20	54 GONZALEZ	SAGREDO	MATIAS IGNACIO	2017	163	207 MORENO	BASAURE	NATALIA VALENTINA
21	5 MIRANDA	MARCEL	CAMILO ANTONIO	2017	164	169 MOYA	ROSAS	VERÓNICA PAZ
22	71 PUELMA	CABALLERO	RODRIGO ANDRÉS	2017	165	162 PEÑA	VEAS	OCTAVIO IGNACIO
23	60 RAMÍREZ	ROBLEDO	NICOLÁS ANTONIO	2017	166	31 SAAVEDRA	PAYACAN	ISABEL MARGARITA
24	70 SEPÚLVEDA	NEAGO	JOSEFINA BELÉN	2017	167	155 SENEL	ZAMORA	CHRISTIAN FELIPE
25	69 SEPÚLVEDA	ROJAS	CRISTÓBAL OCTAVIO	2017	168	13 SEPÚLVEDA	LEMUS	ERNESTO
26	91 TRONCOSO	CERDA	MARÍA ESPERANZA	2017	169	81 SILVA	JIMÉNEZ	ALONSO MARCELO
27	57 VALDIVIA	RETAMAL	PABLO ANDRÉS	2017	170	53 SILVA	PÉREZ	CRISTÓBAL JOAQUÍN
28	95 ALEGRIA	LANAS	KATALINA PÍA ANTONELLA	2018	171	167 SMITH	CASTILLO	VANESSA ALEXANDRA
29	65 ARAYA	JARA	HÉCTOR BENJAMÍN	2018	172	90 TAMPE	MEDINA	JAVIER ALONSO
30	22 ARRAGADA	BONILLA	BÁRBARA KARINA	2018	173	183 TOLEDO	RODRÍGUEZ	ANTONIA SOFÍA
31	121 BAEZ	DUPRE	ANTONIA RAYEN	2018	174	103 TRINCADO	MOYA	JOSELINE SCARLET
32	217 BUILES	TABORDA	ANA MARIA	2018	175	205 AROS	CARVALLO	ISIDORA MELISA
33	3 CLAVEROL	MARTINEZ	PEDRO ANTONIO	2018	176	141 ARRIAGADA	LINEROS	JAVIERA BELEN
34	72 COLOMA	SOLÍS	CATALINA JAZMÍN	2018	177	193 ASENCO	MORA	KRISHNA PAULINA
35	76 ESPINOZA	ARRIOLA	EDNA ISABELLA	2018	178	199 BAEZ	CORREA	AMANDA CRISTINA
36	68 GARRIDO	ALARCÓN	PAULINA JAVIERA	2018	179	216 BALLERINO	IBÁÑEZ	SOFÍA VALENTINA
37	87 HALFER	DÉREZ	CATALINA AURORA	2018	180	192 BRAVO	LÓPEZ	PILAR CATALINA
					181	210 CAMPUSANO	RODRÍGUEZ	MARTÍN ANTONIO

- El marco muestral es **CLAVE**: la muestra recogida sólo podrá considerarse representativa de la población que está comprendida en el marco muestral
- Definimos entonces:
 - **Población/Universo**: agregado de elementos respecto del cual se recaba información
 - **Población objetivo**: es una parte de la población, excluye elementos de difícil acceso o que son muy caros de encuestar
 - **Población marco**: es una parte de la población objetivo; se excluyen los elementos que no son encuestados por marco insuficiente o falta de cobertura
 - **Población de encuesta**: es una parte de la población marco; se excluyen los distintos tipos de no respuesta

Datos de la muestra

UNIVERSO QUE SE INCLUYE

Población de 18 años y más (urbana y rural) residente a lo largo de todo el país (a excepción de Isla de Pascua, Juan Fernández y la Antártica).

COBERTURA DEL UNIVERSO

Toda la población de 18 años y más en base al Censo de 2017 (a excepción de Isla de Pascua, Juan Fernández y la Antártica).

MUESTRA

1.355 personas fueron entrevistadas en sus hogares, en **126 comunas** del país.

El método de muestreo fue estratificado (por región y zona urbana/rural), aleatorio y probabilístico en cada una de sus tres etapas (manzana-hogar-entrevistado). No se utilizaron reemplazos.

El nivel de respuesta fue de 61,1% lo que corresponde a las 1.355 entrevistas completas sobre la muestra original de 2.216 entrevistas.

NIVEL DE PRECISIÓN

El error muestral se estima en ± 2,9% considerando varianza máxima, un nivel de confianza **95%** y un efecto del diseño estimado de 1,2.

- Población/Universo: total: Población con derecho a voto (18 años o más)
- Población objetivo: Población – Población difícil acceso o muy cara (- Isla de Pascua, Juan Fernández y la Antártica)
- Población marco: Población objetivo – la que no está en el marco o falta de cobertura: considerando la construcción de un marco muestral
- Población encuesta: Población marco – población que no responde: menos quienes no responden (¿tienen alguna particularidad o se distribuyen de forma aleatoria?)

¿Cómo se calcula el tamaño de la muestra?

Disponibilidad del marco muestral

Tamaño de la población y varianza poblacional

Nivel de precisión deseado
(margen de error y nivel de confianza de la estimación)

Complejidad de los análisis estadísticos

Tiempo y recursos

Procedimiento de muestreo seleccionado

Algunos conceptos claves

- **Intervalo de Confianza (IC):** Rango de valores que tiene una cierta probabilidad conocida de contener el verdadero valor del parámetro de la población (ej: {5,6; 6,7} o {35%; 52%}): valor se mueve entre X e Y.
- **Nivel de Confianza:** Es la probabilidad $1 - \alpha$, que es la proporción de veces que el intervalo de confianza realmente contiene el parámetro de población. Lo elige el/la investigador/a, generalmente 90%, 95% o 99%: con cuánta confianza mi muestra está en el IC?
- **El error máximo de estimación:** Nos indica el mayor error que podemos cometer por predecir la proporción poblacional con una proporción muestral (ej: 2%): cuánto error estoy dispuesto admitir?

Datos de la muestra

UNIVERSO QUE SE INCLUYE

Población de 18 años y más (urbana y rural) residente a lo largo de todo el país (a excepción de Isla de Pascua, Juan Fernández y la Antártica).

COBERTURA DEL UNIVERSO

Toda la población de 18 años y más en base al Censo de 2017 (a excepción de Isla de Pascua, Juan Fernández y la Antártica).

MUESTRA

1.355 personas fueron entrevistadas en sus hogares, en **126 comunas** del país.

El método de muestreo fue estratificado (por región y zona urbana/rural), aleatorio y probabilístico en cada una de sus tres etapas (manzana-hogar-entrevistado). No se utilizaron reemplazos.

El nivel de respuesta fue de **61,1%** lo que corresponde a las 1.355 entrevistas completas sobre la muestra original de 2.216 entrevistas.

NIVEL DE PRECISIÓN

El error muestral se estima en **± 2,9%**, considerando varianza máxima, **un nivel de confianza 95%** y un efecto del diseño estimado de **1,2**.

- a) Nivel de confianza de 95%
- b) Error de estimación +- 2,9%
- c) Varianza máxima: 0,5
- d) Efecto diseño: 1,2

¿Cuántos **casos** son necesarios?

- **Cálculo del tamaño de la muestra mediante MAS (muestreo aleatorio simple)**
- La fórmula incorpora el tamaño de la población, la varianza, el error máximo de estimación y el nivel de confianza deseado.....

- Tamaño de muestra de proporciones en MAS

Poblaciones finitas

$$n = \frac{z^2 \cdot PQ \cdot N}{(N - 1)e^2 + z^2PQ}$$

- N: Tamaño de población

• PQ: Variación poblacional
(0,5 * 0,5; máxima)

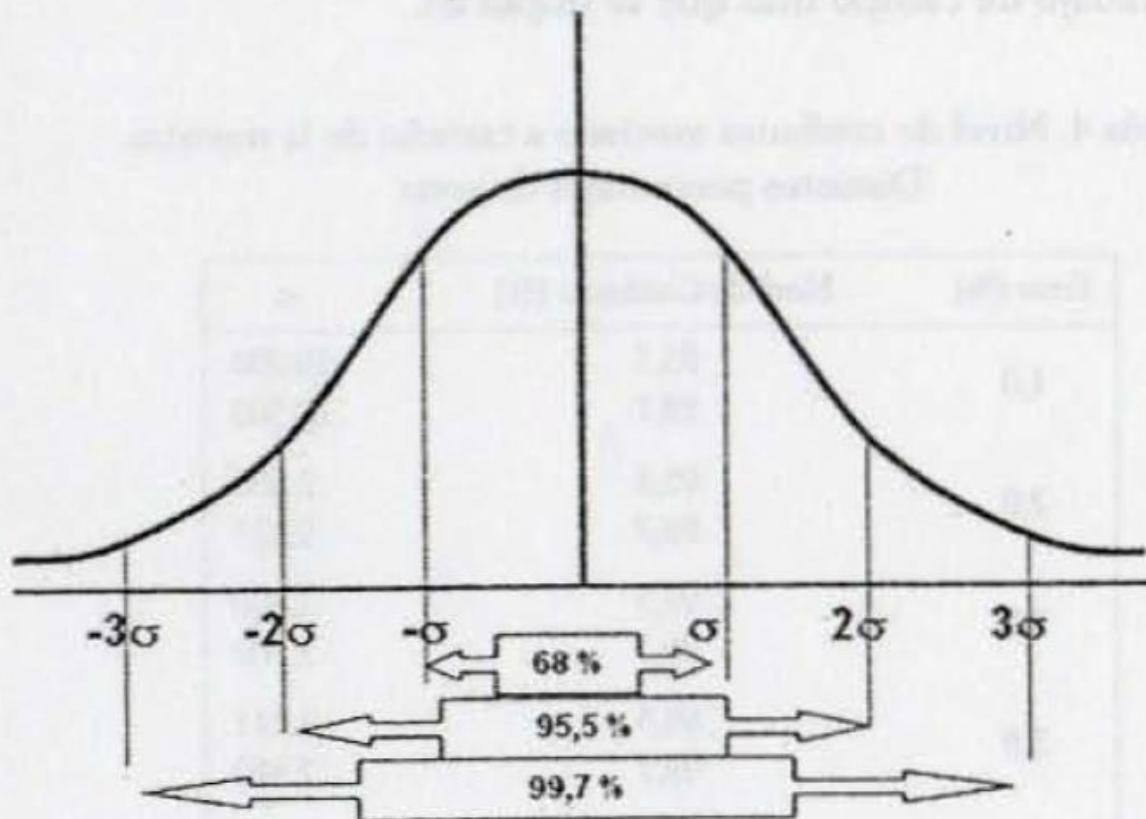
• e: Error máximo admisible
(mitad de IC)

Poblaciones infinitas

$$n = \frac{z^2 \cdot PQ}{e^2}$$

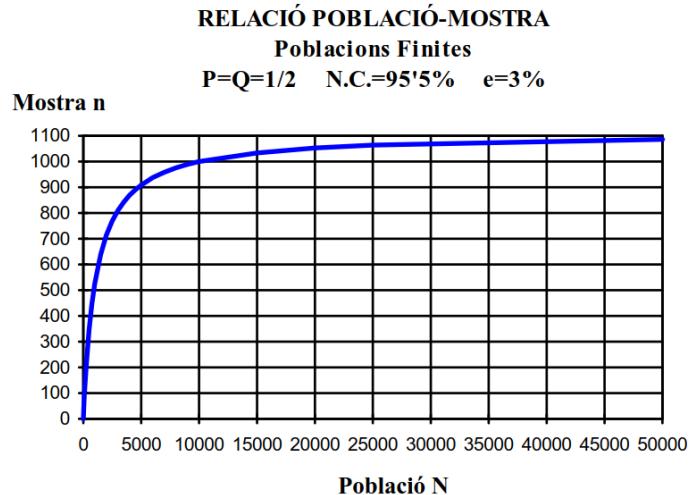
Población/Universo mayor a 10.000

puntuación z (Desviaciones estándar)	valor P (Probabilidad)	Nivel de confianza
< -1,65 o > +1,65	< 0,10	90%
< -1,96 o > +1,96	< 0,05	95%
< -2,58 o > +2,58	< 0,01	99%



$$n = \frac{z^2 \cdot PQ \cdot N}{N \cdot e^2 + z^2 \cdot PQ} = \frac{4 \cdot 0,25 \cdot N}{0,3^2 \cdot N + 4 \cdot 0,25} = \frac{N}{0,0009 \cdot N + 1}$$

N	n
100	92
250	204
500	345
750	448
1000	526
1500	638
2000	714
2500	769
3000	811
3500	843
4000	870
4500	891
5000	909
6000	938
7000	959
8000	976
9000	989
10000	1000
15000	1034
20000	1053
25000	1064
50000	1087



Ejercicio

Formulas Muestra

Poblaciones finitas

$$n = \frac{z^2 \cdot PQ \cdot N}{(N - 1)e^2 + z^2PQ}$$

Z	1,96
e	3,00
N	1.500
n	775
P	50
Q=1-P	50

n? 623,77

e? 2,45

Poblaciones infinitas

$$n = \frac{z^2 \cdot PQ}{e^2}$$

Z	1,96
e	3
n	1.071
P	50
Q=1-P	50

n? 1067,11

e? 2,9945

Considerando muestras infinitas

- ¿Cuántos casos se necesitan para una muestra aleatoria simple con un error de 5% y un nivel de confianza de 90% y una varianza máxima?
- ¿Cuántos casos se necesitan para una muestra aleatoria simple con un error de 2% y un nivel de confianza de 95% y una varianza máxima?

puntuación z (Desviaciones estándar)	valor P (Probabilidad)	Nivel de confianza
< -1,65 o > +1,65	< 0,10	90%
< -1,96 o > +1,96	< 0,05	95%
< -2,58 o > +2,58	< 0,01	99%

Tipos de muestreos

Muestreo Probabilístico

- Todos los elementos de la población tienen una probabilidad conocida y distinta de 0 de pertenecer a la muestra.
- Se sustenta en el principio de selección aleatoria.
- Permite la inferencia
- Es más caro y más complejo, pero más Preciso

Muestreo No Probabilístico

- Hay una desigual probabilidad de las unidades de la población de formar parte en la muestra
- No prioriza la aleatorización
- Se usa por conveniencia (tiempo, recursos) o cuando la inferencia no es lo central
- Es más barato y simple, pero menos preciso

Muestreos Probabilísticos



Muestreo Aleatorio Simple (MAS)

- La selección se realiza de un listado de la población, asignándole **igual probabilidad a cada elemento**.
- Se caracteriza por:
 - Cada unidad tiene una **probabilidad igual** de participar en la selección
 - La **selección es totalmente aleatoria**
 - Todas las posibles **muestras** tienen igual probabilidad de ser elegidas



Muestreo Aleatorio Simple

 QuestionPro

Selección Aleatoria

- Por lo general la selección aleatoria se hace **computacionalmente**. También existen tablas de números aleatorios.
- Como el **MAS exige un listado exhaustivo**, es **imposible** utilizarlo en poblaciones de magnitud elevada. Por ejemplo, es imposible hacer un MAS en todo Chile.

Listado de
estudiantes de
Antropología
Universidad
Alberto Hurtado:
220 personas

Nº	Apellido Paterno	Apellido Materno	Nombres	Año Admisión				
2	4 TAPIA	VALLADARES	TAMARA JACQUELINE	2011	145	79 AGUILAR	TRONCOSO	SOFÍA VALENTINA
3	6 HERNANDEZ	HERNANDEZ	ALEJANDRO SALVADOR	2012	146	30 ALDANA	KURTE	CRISTOBAL FRANCISCO
4	11 COO	SANHUEZA	SOFIA INÉS ADRIANA	2014	147	7 ALVAREZ	REYES	MATIAS FERNANDO
5	1 OLIVA	CORNEJO	VALENTINA SOLEDAD	2014	148	129 ASTUDILLO	RAMÍREZ	NAYARETH ROCÍO IGNACIA
6	18 URRIOLA	ABALLAI	CATALINA ANDREA	2014	149	204 BUSTOS	PALMA	JAVIERA VIVIANA
7	23 PEÑA	SOTO	JAVIERA CONSTANZA	2015	150	12 CONCHA	CAICEDO	GABRIEL ROBERTO
8	29 CAQUISANE	TAPIA	JAVIERA FERNANDA	2016	151	144 CORNEJO	ROJAS	PABLO TOMÁS ARIEL
9	39 LAGOS	CARTES	CAROLINA ANDREA	2016	152	185 ESCAFI	BASULTO	SAMANTHA MICHELLE
10	9 LAS HERAS	TAPIA	IVAN DAVID	2016	153	104 ESPINOZA	MORALES	MARÍA JOSÉ
11	34 OPORTO	MORAN	NAYARET CATALINA	2016	154	50 FIGUEROA	ZAMORANO	CATALINA JAVIERA
12	48 ORTIZ	LATAPIAT	MARTINA ANGELICA	2016	155	219 FOGARTY	CHRYSTAL	SHEILA MAUREEN
13	26 PEREZ	JIMENEZ	MARIA IGNACIA	2016	156	181 FREDES	MORA	FERNANDA IGNACIA
14	16 VERA	REIMÁN	CLAUDIA ANTONIA	2016	157	94 GONZÁLEZ	NAVARRETE	ESTEFANO FABIÁN
15	40 ARANZAES	HERRÁN	CAROLINA ANTONIA	2017	158	175 HENRÍQUEZ	BRIONES	PAULINA IGNACIA
16	20 BASTÍAS	ÁLVAREZ	FELIPE ANDRÉS	2017	159	195 HERRERA	CATALAN	JAVIERA VIOLETA
17	15 CAUCHUPAN	GAJARDO	JAVIERA VALENTINA PAZ	2017	160	184 MATTE	DÍAZ	ELISA
18	10 COX	VERDUGO	MATIAS IGNACIO	2017	161	126 MAYORA	PICKER	NICOLAS
19	37 FIERRO	ESCOBAR	ANDRES SIMON	2017	162	132 MELÉNDEZ	HENRÍQUEZ	JOAQUÍN ANDRÉS
20	54 GONZALEZ	SAGREDO	MATIAS IGNACIO	2017	163	207 MORENO	BASAURE	NATALIA VALENTINA
21	5 MIRANDA	MARCEL	CAMILO ANTONIO	2017	164	169 MOYA	ROSAS	VERÓNICA PAZ
22	71 PUELMA	CABALLERO	RODRIGO ANDRÉS	2017	165	162 PEÑA	VEAS	OCTAVIO IGNACIO
23	60 RAMÍREZ	ROBLEDO	NICOLÁS ANTONIO	2017	166	31 SAAVEDRA	PAYACAN	ISABEL MARGARITA
24	70 SEPÚLVEDA	NEAGO	JOSEFINA BELÉN	2017	167	155 SENEL	ZAMORA	CHRISTIAN FELIPE
25	69 SEPÚLVEDA	ROJAS	CRISTÓBAL OCTAVIO	2017	168	13 SEPÚLVEDA	LEMUS	ERNESTO
26	91 TRONCOSO	CERDA	MARÍA ESPERANZA	2017	169	81 SILVA	JIMÉNEZ	ALONSO MARCELO
27	57 VALDIVIA	RETAMAL	PABLO ANDRÉS	2017	170	53 SILVA	PÉREZ	CRISTÓBAL JOAQUÍN
28	95 ALEGRIA	LANAS	KATALINA PÍA ANTONELLA	2018	171	167 SMITH	CASTILLO	VANESSA ALEXANDRA
29	65 ARAYA	JARA	HÉCTOR BENJAMÍN	2018	172	90 TAMPE	MEDINA	JAVIER ALONSO
30	22 ARRAGADA	BONILLA	BÁRBARA KARINA	2018	173	183 TOLEDO	RODRÍGUEZ	ANTONIA SOFÍA
31	121 BAEZ	DUPRE	ANTONIA RAYEN	2018	174	103 TRINCADO	MOYA	JOSELINE SCARLET
32	217 BUILES	TABORDA	ANA MARIA	2018	175	205 AROS	CARVALLO	ISIDORA MELISA
33	3 CLAVEROL	MARTINEZ	PEDRO ANTONIO	2018	176	141 ARRIAGADA	LINEROS	JAVIERA BELEN
34	72 COLOMA	SOLÍS	CATALINA JAZMÍN	2018	177	193 ASENCO	MORA	KRISHNA PAULINA
35	76 ESPINOZA	ARRIOLA	EDNA ISABELLA	2018	178	199 BAEZ	CORREA	AMANDA CRISTINA
36	68 GARRIDO	ALARCÓN	PAULINA JAVIERA	2018	179	216 BALLERINO	IBÁÑEZ	SOFÍA VALENTINA
37	87 HALFER	DÉREZ	CATALINA AURORA	2018	180	192 BRAVO	LÓPEZ	PILAR CATALINA
					181	210 CAMPUSANO	RODRÍGUEZ	MARTÍN ANTONIO

Ventajas y desventajas del MAS

VENTAJAS

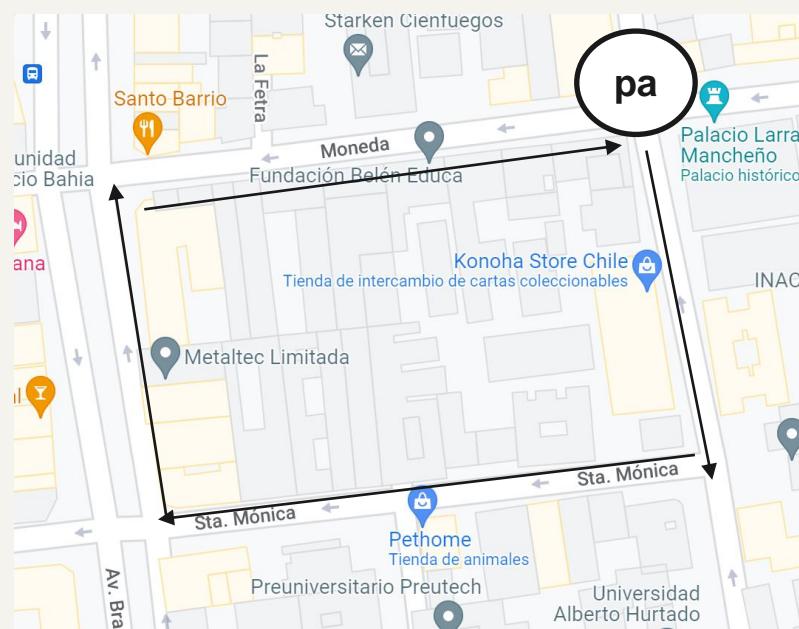
- Facilidad de los cálculos estadísticos
- Elevada probabilidad de ser representativo

DESVENTAJAS

- Requiere listar y enumerar a las unidades de la población
- Resulta monótono arduo en muestras grandes
- La dispersión de la muestra hace que los costos sean elevados

Muestreo Sistemático

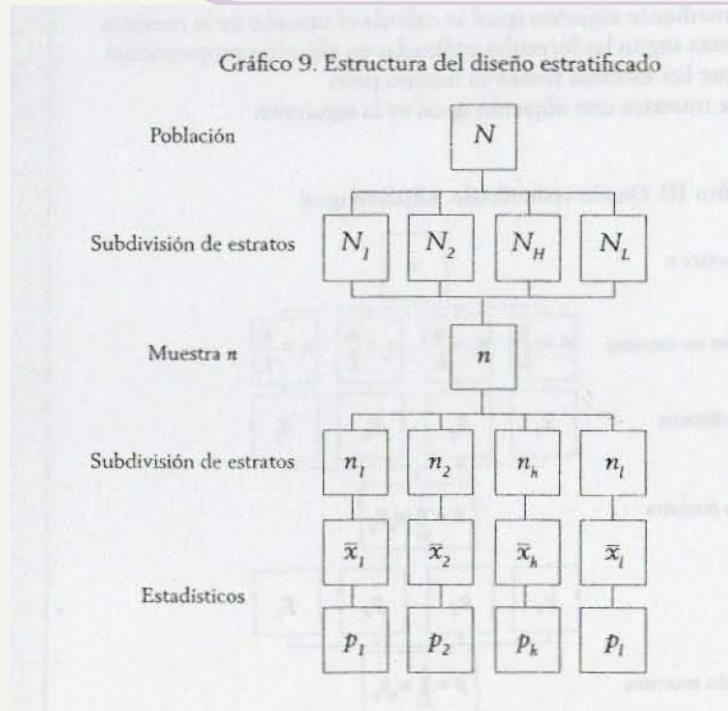
- Es una variante del MAS. Se elige un punto de arranque aleatorio y se seleccionan casos en forma sucesiva según un intervalo numérico convencional.
- Sólo la primera unidad de la muestra se elige al azar
- Los restantes elementos de la muestra se eligen sumando el coeficiente de elevación que es igual a N/n , hasta llegar al n requerido
- En trabajo en terreno es mucho más práctico que un MAS



	Nº	Apellido Paterno	Apellido Materno	Nombres	Año Admisión	total/cantidad de casos	salto de
2	4	TAPIA	VALLADARES	TAMARA JACQUELINE	2011	220/44	5
3	6	HERNANDEZ	HERNANDEZ	ALEJANDRO SALVADOR	2012		
4	11	COO	SANHUEZA	SOFIA INÉS ADRIANA	2014		
5	1	OLIVA	CORNEJO	VALENTINA SOLEDAD	2014		
6	18	URRIOLA	ABALLAI	CATALINA ANDREA	2014		
7	23	PEÑA	SOTO	JAVIERA CONSTANZA	2015		
8	29	CAQUISANE	TAPIA	JAVIERA FERNANDA	2016		
9	39	LAGOS	CARTES	CAROLINA ANDREA	2016		
10	9	LAS HERAS	TAPIA	IVAN DAVID	2016		
11	34	OPORTO	MORAN	NAYARET CATALINA	2016		
12	48	ORTIZ	LATAPIAT	MARTINA ANGELICA	2016		
13	26	PEREZ	JIMENEZ	MARIA IGNACIA	2016		
14	16	VERA	REIMÁN	CLAUDIA ANTONIA	2016		
15	40	ARANZAES	HERRÁN	CAROLINA ANTONIA	2017		
16	20	BASTÍAS	ÁLVAREZ	FELIPE ANDRÉS	2017		
17	15	CAUCHUPAN	GAJARDO	JAVIERA VALENTINA PAZ	2017		
18	10	COX	VERDUGO	MATIAS IGNACIO	2017		
19	37	FIERRO	ESCOBAR	ANDRES SIMON	2017		
20	54	GONZALEZ	SAGREDO	MATIAS IGNACIO	2017		
21	5	MIRANDA	MARCEL	CAMILO ANTONIO	2017		
22	71	PUELMA	CABALLERO	RODRIGO ANDRÉS	2017		
23	60	RAMÍREZ	ROBLEDO	NICOLÁS ANTONIO	2017		
24	70	SEPÚLVEDA	NEAGO	JOSEFINA BELÉN	2017		
25	69	SEPÚLVEDA	ROJAS	CRISTÓBAL OCTAVIO	2017		
26	91	TRONCOSO	CERDA	MARÍA ESPERANZA	2017		
27	57	VALDIVIA	RETAMAL	PABLO ANDRÉS	2017		
28	95	ALEGRIA	LANAS	KATALINA PÍA ANTONELLA	2018		
29	65	ARAYA	JARA	HÉCTOR BENJAMÍN	2018		
30	22	ARRIAGADA	BONILLA	BÁRBARA KARINA	2018		
31	121	BAEZ	DUPRE	ANTONIA RAYEN	2018		
32	217	BUILES	TABORDA	ANA MARIA	2018		
33	3	CLAVEROL	MARTINEZ	PEDRO ANTONIO	2018		
34	72	COLOMA	SOLÍS	CATALINA JAZMÍN	2018		
35	76	ESPINOZA	ARRIOLA	EDNA ISABELLA	2018		
36	68	GARRIDO	ALARCÓN	PAULINA JAVIERA	2018		
37	87	HAUER	PÉREZ	CATALINA AURORA	2018		

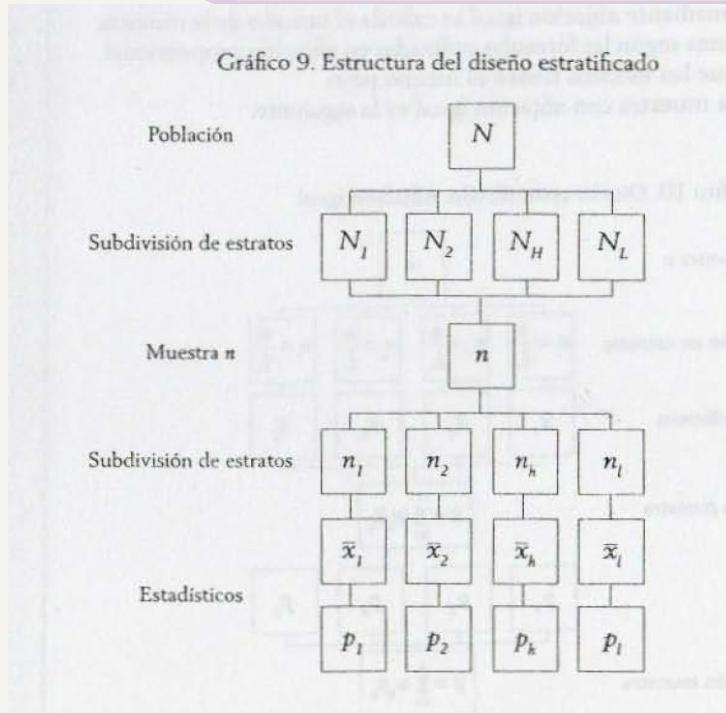
Muestreo Estratificado

- Se caracteriza por la utilización de información auxiliar para mejorar la eficiencia en la selección de elementos y mejorar la precisión de las estimaciones.
- La información auxiliar corresponde a **variables de estratificación** que dividen a la población en estratos.
- Las variables de estratificación más utilizadas son edad, sexo, clase social, ocupación. Esto se debe a que tienden a estar recogidas en los marcos muestrales (como el CENSO)



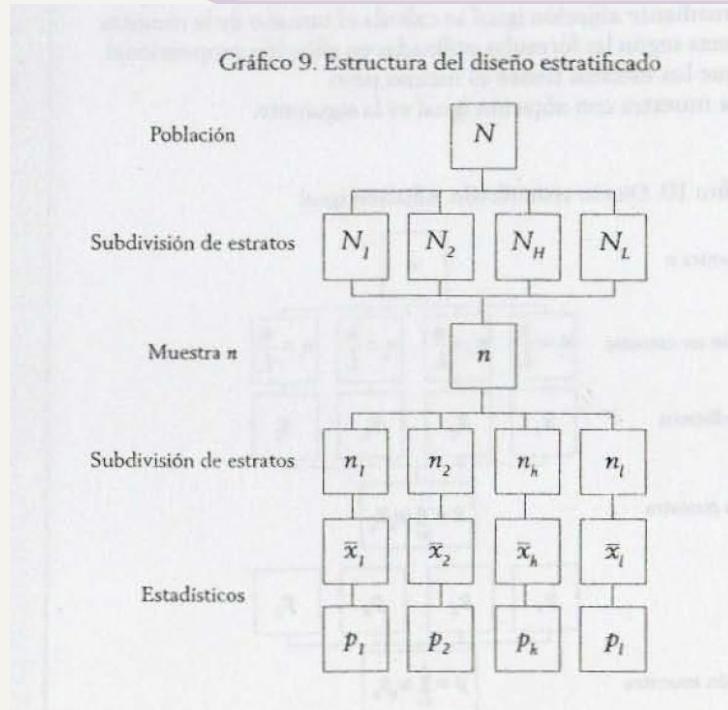
Muestreo Estratificado

- ¿Qué es un estrato? → Los estratos son **grupos homogéneos de elementos** por lo que no es preciso seleccionar un número elevado de casos para representar al estrato.
- Mientras más **homogéneo** sea el estrato, más precisa será la estimación.
 - Las varianza entre los estratos es máxima y al interior de cada estrato es mínima
- El fin es lograr que el estrato esté representado en la muestra.
- Logrará estimaciones más precisas, siempre que:
 - Las variables de estratificación están relacionadas con los objetivos de la investigación.
 - De no ser así, no tendría sentido (se logaría una precisión similar a si no se hubiera estratificado).



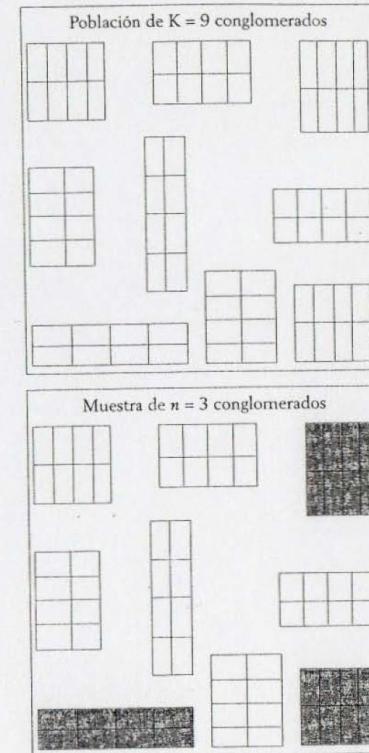
Muestreo Estratificado

- Afijación proporcional al tamaño
- Afijación no proporcional

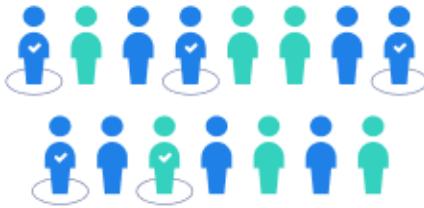


Muestreo por Conglomerados

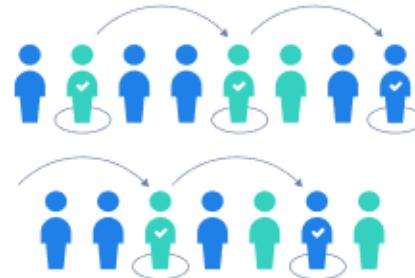
- Se caracteriza por constituir unidades de muestreo compuestas por un agregado diverso de elementos.
- ¿Qué es un conglomerado? → conjunto de elementos con valores diferentes entre sí en la variable medida.
- Diferencias con el ME:
 - La lógica es inversa. El error disminuye en la medida que aumenta la heterogeneidad de cada uno de los conglomerados
 - Se extrae una muestra aleatoria de conglomerados (no una muestra al interior de cada estrato)
 - La unidad de muestreo es el conglomerado y no el individuo



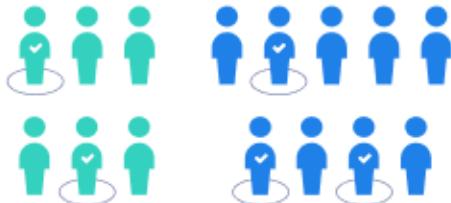
Simple random sample



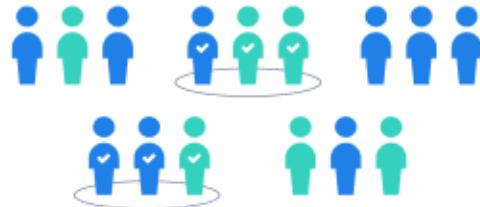
Systematic sample



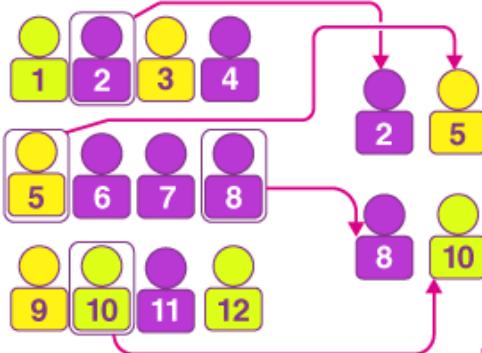
Stratified sample



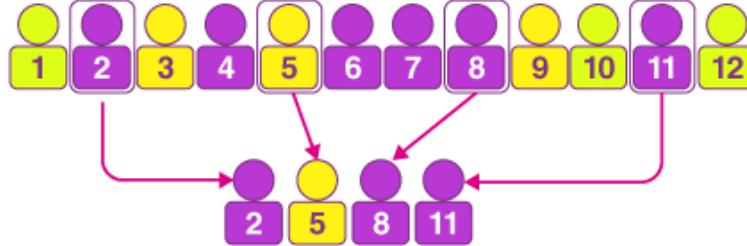
Cluster sample



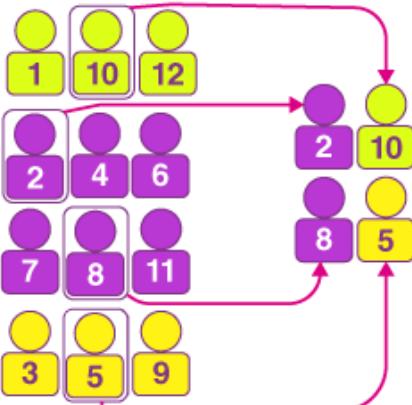
Simple Random Sampling



Systematic Sampling

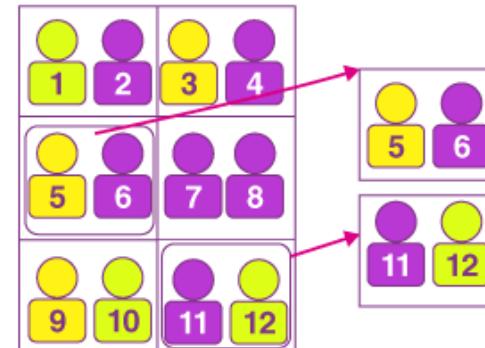


Stratified Sampling



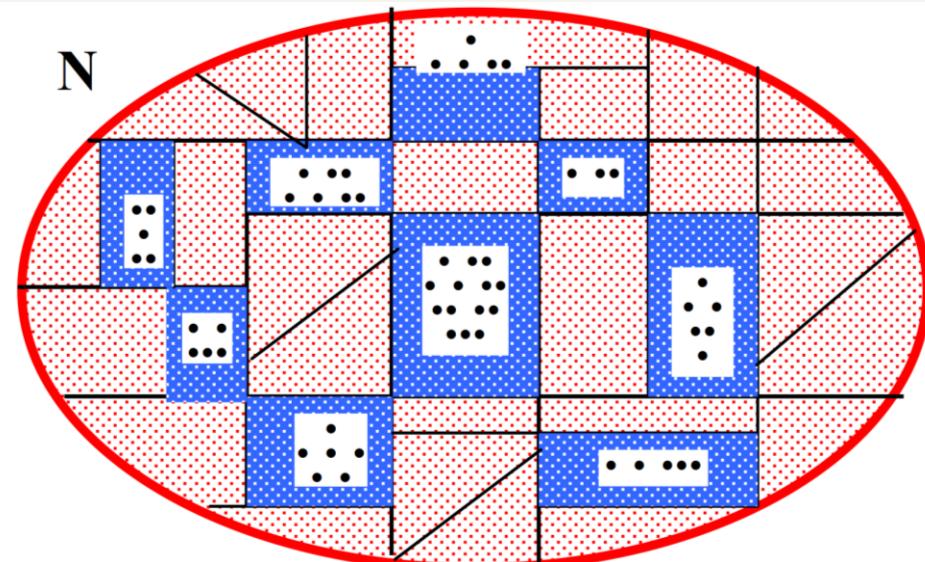
Probability sampling Methods

Clustered Sampling



Muestreo Polietápico

- El MC da paso comúnmente a los muestreos polietápicos....
 - Son diseños complejos caracterizados por la selección de muestras en etapas sucesivas.
- Cada etapa requiere su propio marco muestral. Hay así unidades primarias de muestreo (UPM), unidades secundarias de muestreo (USM), etc.
- La versión más sencilla tiene 2 etapas:
 - Se seleccionan los conglomerados (unidades de muestreo primarias)
 - Se elige mediante MAS a los miembros de la población a observar, al interior de los conglomerados previamente seleccionados
- Es el más típico muestreo nacional → por ejemplo, muestreo polietápico, estratificado, por conglomerados.



Datos de la muestra

UNIVERSO QUE SE INCLUYE

Población de 18 años y más (urbana y rural) residente a lo largo de todo el país (a excepción de Isla de Pascua, Juan Fernández y la Antártica).

COBERTURA DEL UNIVERSO

Toda la población de 18 años y más en base al Censo de 2017 (a excepción de Isla de Pascua, Juan Fernández y la Antártica).

MUESTRA

1.355 personas fueron entrevistadas en sus hogares, en **126 comunas** del país.

El método de muestreo fue estratificado (por región y zona urbana/rural), aleatorio y probabilístico en cada una de sus tres etapas (manzana-hogar-entrevistado). No se utilizaron reemplazos.

El nivel de respuesta fue de **61,1%** lo que corresponde a las 1.355 entrevistas completas sobre la muestra original de 2.216 entrevistas.

NIVEL DE PRECISIÓN

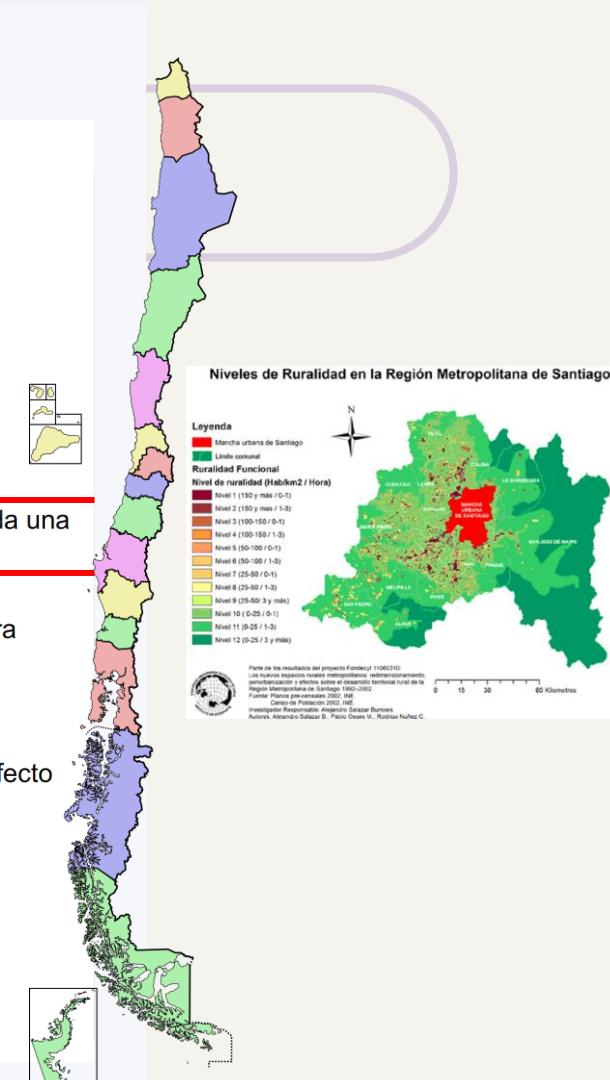
El error muestral se estima en **± 2,9%** considerando varianza máxima, un nivel de confianza **95%** y un efecto del diseño estimado de 1,2.

INSTRUMENTO

Entrevistas cara a cara, basadas en un cuestionario estructurado.

FECHA DE TERRENO

La recolección de datos se efectuó entre el **13 de abril y el 29 de mayo del 2022**.



Niveles de Ruralidad en la Región Metropolitana de Santiago

Leyenda

Mancha urbana de Santiago

Límite comunal

Ruralidad Funcional

Nivel de ruralidad (Hab/km² / Hora)

Nivel 1 (150 y más / 0-1)

Nivel 2 (150 y mas / 1-3)

Nivel 3 (100-150 / 0-1)

Nivel 4 (100-150 / 1-3)

Nivel 5 (50-100 / 0-1)

Nivel 6 (50-100 / 1-3)

Nivel 7 (25-50 / 0-1)

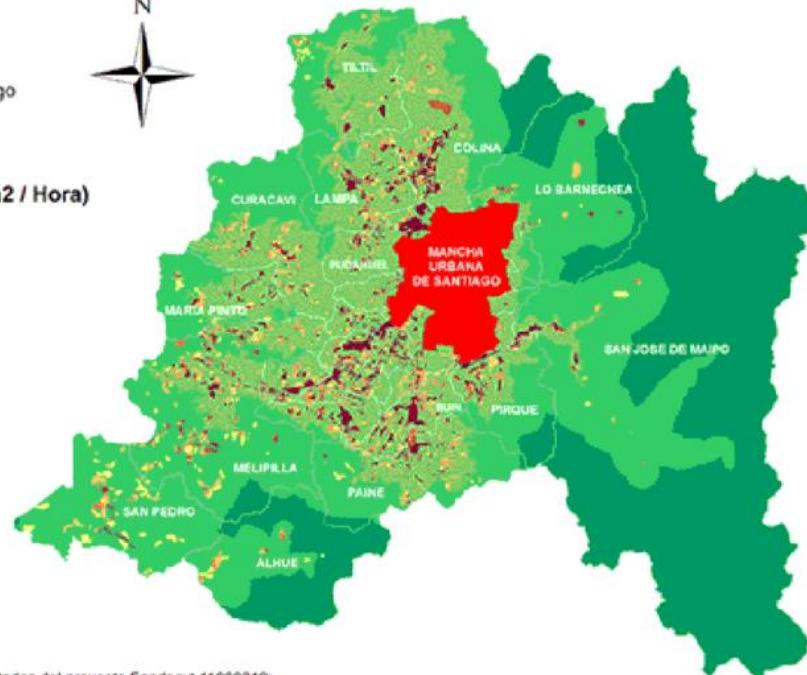
Nivel 8 (25-50 / 1-3)

Nivel 9 (25-50/ 3 y más)

Nivel 10 (0-25 / 0-1)

Nivel 11 (0-25 / 1-3)

Nivel 12 (0-25 / 3 y más)



Parte de los resultados del proyecto Fondecyt 11060310:
Los nuevos espacios rurales metropolitanos: redimensionamiento,
periurbanización y efectos sobre el desarrollo territorial rural de la
Región Metropolitana de Santiago 1992-2002.

Fuente: Planos pre-censales 2002, INE.

Censo de Población 2002, INE.

Investigador Responsable: Alejandro Salazar Burrows

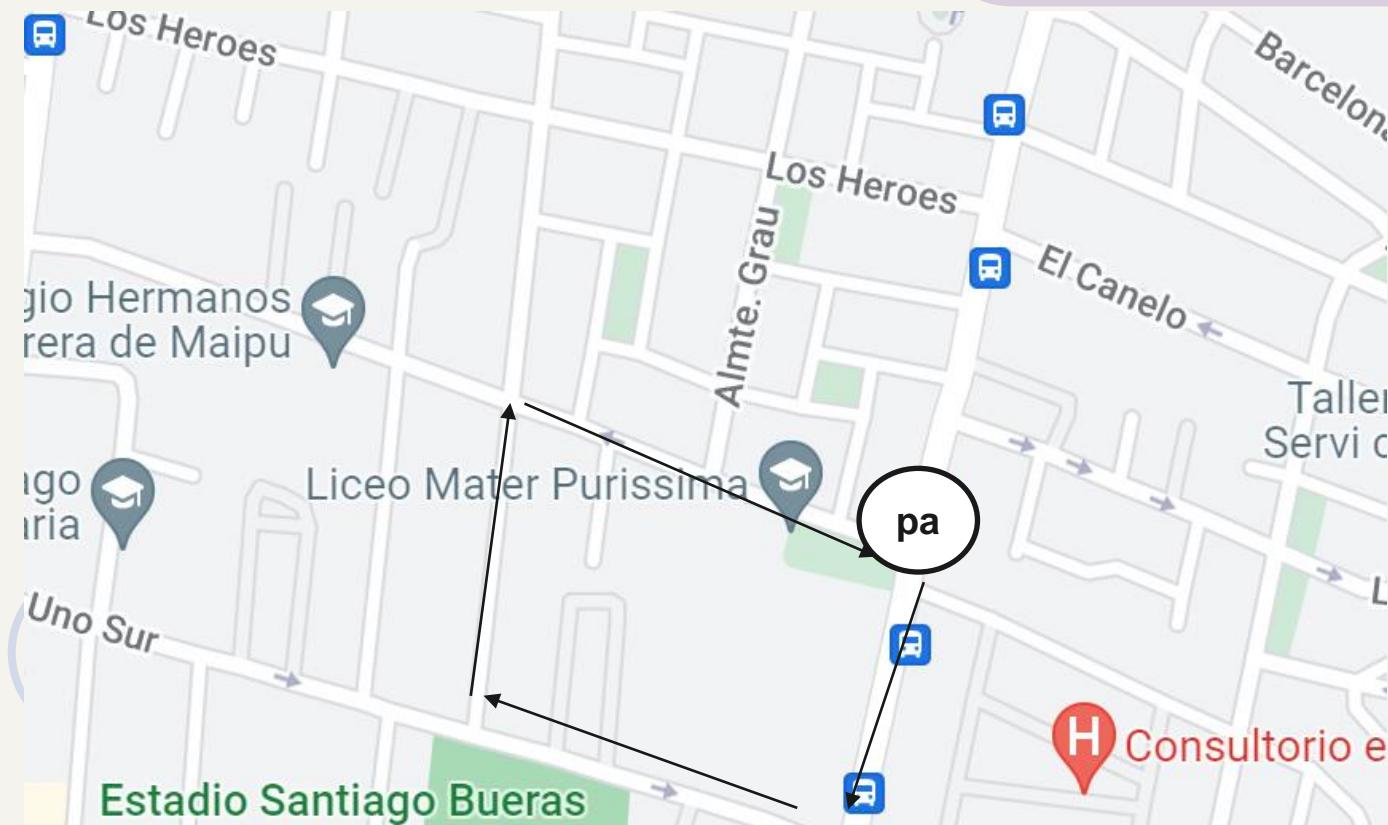
Autores: Alejandro Salazar B., Pablo Osses M., Rodrigo Núñez C.



Objetivo encuesta: Electoral
Paso 1: estratificado región y
urbano/rural

Paso 2: Selecciono Manzanas aleatoriamente

Paso 3: Hago un listado de casas (construyo el marco muestra) y voy seleccionando Casas de una manzana de forma sistemática. Toco el timbre cada 3 casas





Paso 4:

a) Hago un listado de los miembros del hogar mayores de 18:

Juan

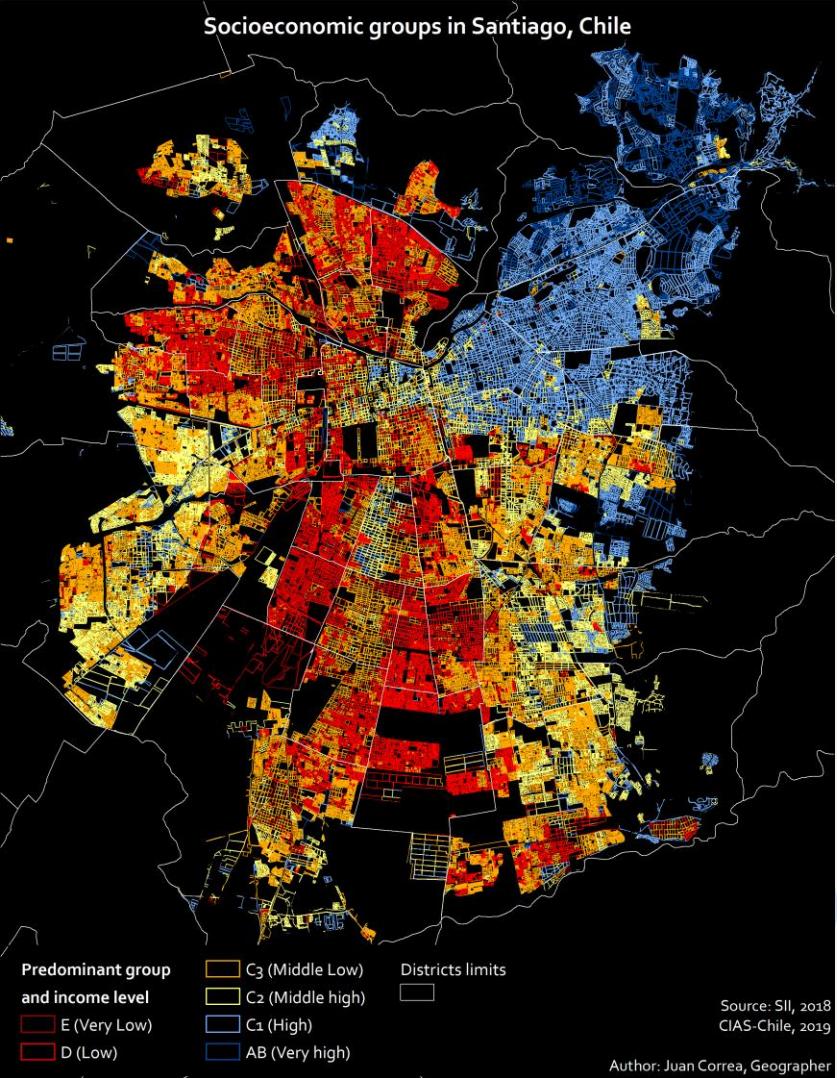
Antonia

Carla

José

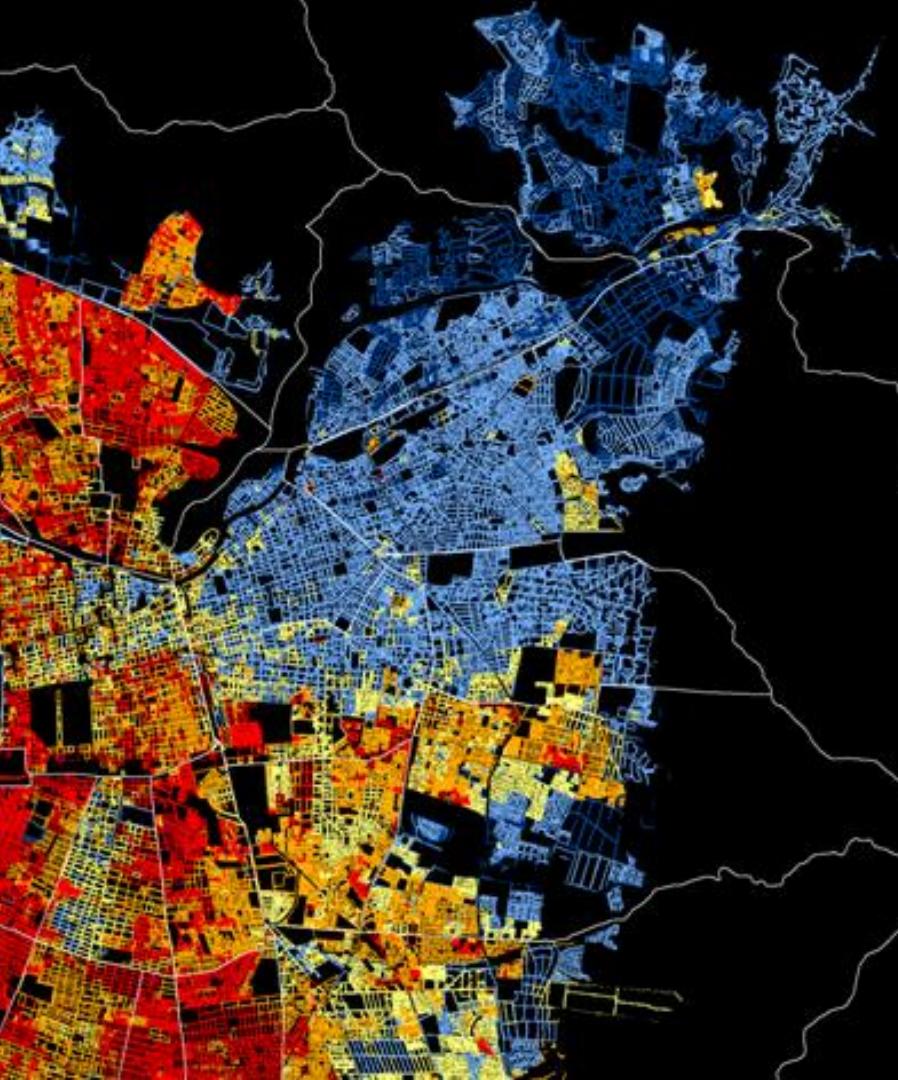
b) luego selecciono de forma aleatoria:
Último que está de cumpleaños

Tabla kish

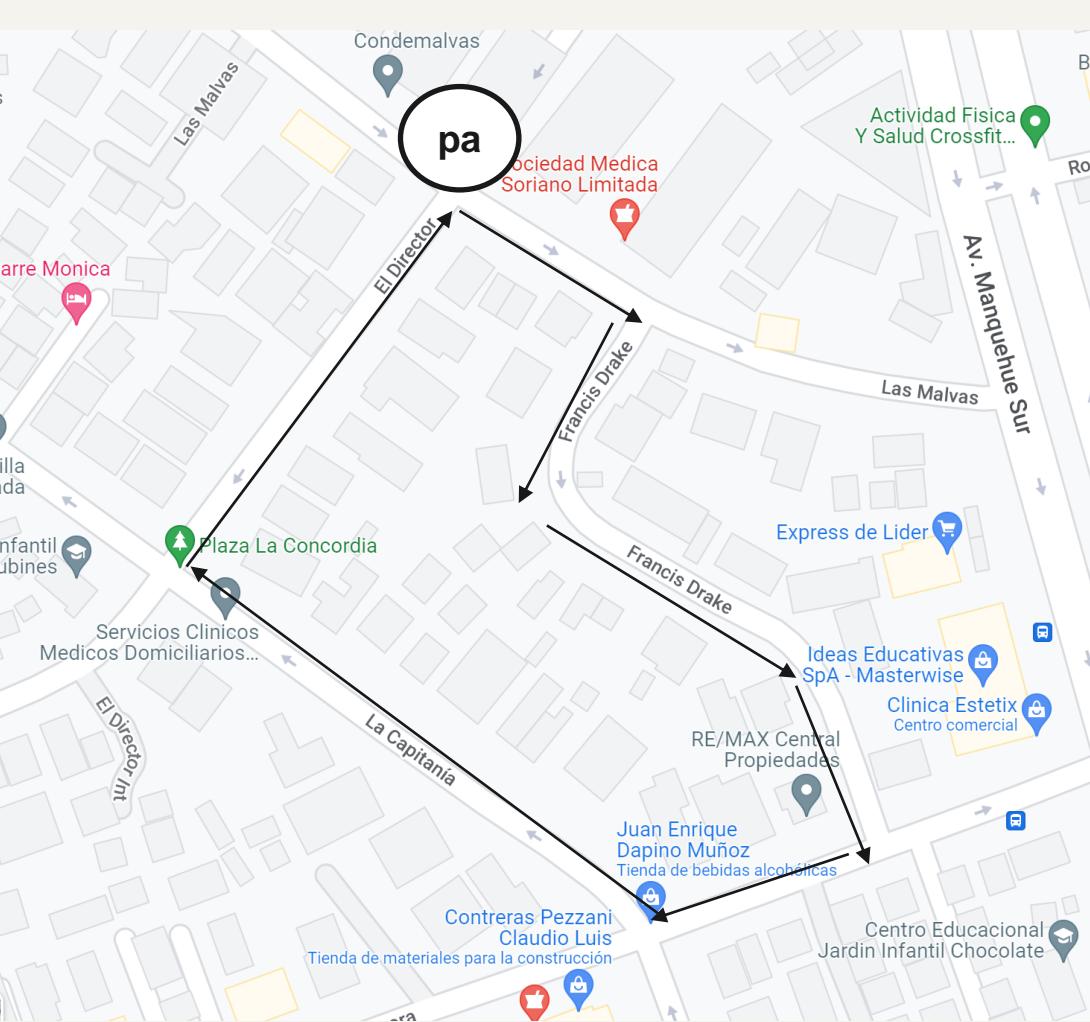


Ejemplo 2:
Objetivo encuesta: Percepciones sobre la
desigualdad social en Santiago

Paso 1: estratifico manzanas por nivel
socioeconómico



Paso 2: Selecciono Manzanas aleatoriamente



Paso 3: Hago un listado de casas
(construyo el marco muestra)
y voy seleccionando Casas de
una manzana de forma
sistemática. Toco el timbre
cada 3 casas

Paso 4:

a) Hago un listado de los miembros del hogar mayores de 18:

- a) José Pablo
- b) María Ignacia
- c) Trini
- d) Jose Miguel

b) luego selecciono de forma aleatoria:

- a) Último que está de cumpleaños
- b) Tabla kish



Muestreos No Probabilísticos



Muestreo por Cuotas

- Es un muestreo no probabilístico muy utilizado. Intenta alcanzar grados mínimos de representatividad de un modo más simple.
- Procedimiento:
 - Se define una matriz con las características básicas de la población que interesa representar y se establece la proporción en que cada una de ellas existe.
 - Luego se determina el n total deseado (en base a criterios de factibilidad y estadísticos) y se distribuye entre cada una de las celdas (cuotas) de manera proporcional
 - Finalmente se le señala a los encuestadores que alcancen la cuota (es decir, que busquen un cierto n de encuestados que cumplan los requisitos de cada cuota)

Ejemplo

- ▶ Se quiere estudiar a una población con las siguientes características:

	Clase Alta	Clase Media	Clase Baja	Total
Hombre	130	180	200	500
Mujer	70	120	300	500
	200	300	500	1000

- ▶ Se calculan las proporciones (el peso de cada grupo):

	Clase Alta	Clase Media	Clase Baja	Total
Hombre	13%	18%	20%	50%
Mujer	7%	12%	30%	50%
	20%	30%	50%	100%

Ejemplo

- ▶ Se decide hacer un muestreo por cuotas con un $n=200$ casos. Se establece el n por cuota (celda) en función de los porcentajes establecidos antes:

	Clase Alta	Clase Media	Clase Baja	Total
Hombre	26	36	40	100
Mujer	14	24	60	100
	40	60	100	200

- ▶ Nótese que los investigadores podrían decidir aumentar el n para tener mínimo 30 casos por celda...

03

Ejercicio en clases

Instrucción

- Revise el marco muestral entregado y trate de identificar que tipo de muestreo podría ser

