



Programa de
Población

Clase 3

Diseño muestral

Introducción a los métodos cuantitativos y
combinados de investigación

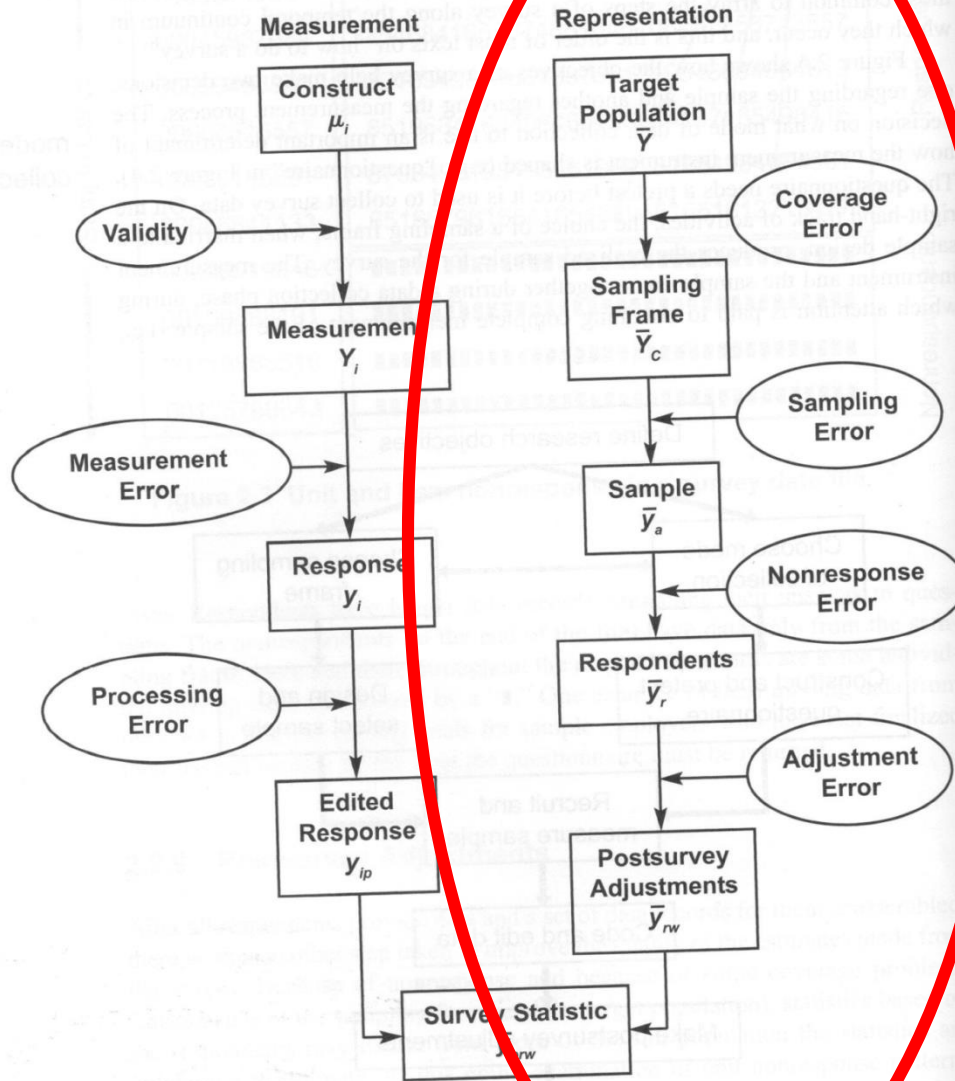
Maestría en Información y Comunicación

Ignacio Pardo

10 de octubre

Objetivos de la clase

- Introducir el concepto de diseño muestral
- Describir las principales variantes de diseño muestral disponibles



Lo que habíamos visto...

- Error de cobertura
- Error muestral
- Error de no respuesta
- Error de ajuste

¿Por qué «funcionan» las muestras?

- ¿En qué sentido funcionan?
 - Datos «superficiales», que se distribuyen en la población de forma previsible
 - ¿Evidencia?
 - Predicción
- Estadística inferencial; probabilidades
- Procesos de selección y estimación
- Base estadística fundamental: teorema central del límite y ley de los grandes números

¿Cuándo «funcionan» las muestras?

- *«Se dice que un estadístico de la muestra es un estimador insesgado para un parámetro de la población si su distribución en el muestreo tiene un valor de media igual al parámetro que va a estimar. En otras palabras, un estimador es insesgado si, en promedio, los valores del estadístico obtenidos del muestreo realmente son iguales al parámetro» (Weinberg & Goldberg, 1982:227)*

¿Cómo «funcionan» las muestras?

- Extracción desde marco muestral: variedad de tipos de muestreo
- ¿Sin marco muestral? Díficil...
 - Ejemplo: ¿Cuántas guarderías hay?
 - Muestreo probabilístico de secciones censales
 - Censo de todas las guarderías de cada sección censal seleccionada
 - ¿Marco *implícito* en *random digit dialing*?

Tipos de muestreo

- **Muestreo probabilístico**
- Muestreo no probabilístico
 - Muestreo intencional o teórico
 - Muestreo de conveniencia
 - Muestreo por bola de nieve
 - Muestreo por cuotas

El muestreo probabilístico...

- ...permite llevar adelante la inferencia estadística entre muestra y población.
- Para eso
 - Aleatoriedad e independencia en la selección de las unidades muestrales...
 - ...y a la hora de la recogida de la información (nuevamente, ver no respuesta y criterios de sustitución)

Población total y desagregación

- ¿A qué nivel de desagregación nos permite nuestra encuesta inferir conclusiones?
 - Debemos decidir si nos alcanza con conclusiones para toda la población o si queremos representatividad en ciertas regiones o subpoblaciones, en cuyo caso debemos cargar la muestra de forma específica
- Diseños muestrales no probabilísticos, pueden «funcionar» como tales (estrictamente, no permiten calcular errores de muestreo).
 - Es el caso del muestreo por cuotas
 - El diseño de la muestra y la selección pueden ser probabilísticos, y que el muestreo por cuotas se dé en la última parte del proceso

Tipos de muestreos básicos

- Los más importantes son:
 - Muestreo aleatorio simple
 - Muestreo aleatorio sistemático
 - Muestreo aleatorio estratificado
 - Muestreo aleatorio por conglomerados
 - La combinación de algunos de estos tipos
- El tipo de muestreo repercute en el proceso y en los errores muestrales

Tamaño de la muestra

- Objetivos de la investigación
- Marco muestral
- ¡Limitaciones económicas!
- Cuanto más conozcamos el marco muestral, mejores muestras extraeremos.
 - Por ejemplo, en las zonas urbanas podría haber una mayor tasa de muestreo que en las rurales, por haber mayor heterogeneidad y por tanto mayor varianza (en aquello que queremos medir)

Tamaño de la muestra

- ¿Qué variables se estimarán?
- ¿Con qué estimadores?
 - Medias, totales, proporciones, coeficientes de modelos estadísticos...
- ¿Con qué procedimientos?
 - (¿intervalos de confianza?)
- Características de la población: distribución de atributos de interés

Muestreo Aleatorio Simple

- Selección aleatoria de las unidades, de una sola vez y sin reemplazo
 - Tablas de números aleatorios, bolillero, métodos informáticos...
 - Se trata de asegurar la equiprobabilidad
- Investigaciones sobre poblaciones pequeñas, específicas y bien diferenciadas (médicos montevideanos, por ej.)
- El cálculo de los errores de muestreo se basa en el M.A.S. (se trata del estándar que se suele invocar, incluso cuando no es este el tipo de muestreo...)
 - El M.A.S. es el patrón para medir la eficacia de los otros tipos de muestreo

Muestreo Aleatorio Sistemático

- Se trata de una variante del MAS.
- Se halla el coeficiente de elevación (N/n) y se elige al azar un número no superior a ese.
- Luego se le suma el coeficiente de elevación al número elegido, todas las veces que sea necesario
 - Por ejemplo, $120000/1500 = 80$
 - Número al azar = 24
 - Unidades seleccionadas = 24, 104 ($80+24$), 184 ($104+24$)...
- Suelen tener menor varianza muestral
 - ¿En qué circunstancias peligra la equiprobabilidad?
Cuando existen patrones en el ordenamiento del marco muestral

Record	Name	Group	
1	Cochran Inc.	Highest	
2	Hu Electronics	Highest	→ Hu Electronics
3	Madow USA	Highest	
4	M.P.H. Bank	Highest	
5	Wolfe & Enix	Highest	
6	Bradburn Corp.	High	
7	Deming Design	High	→ Deming Design
8	HydeBev	High	
9	Sudman Inc.	High	
10	Wallman AG	High	
11	Fuller & Fuller	Medium	
12	Habermann AG	Medium	→ Habermann AG
13	Kalton Group	Medium	
14	Norwood LC	Medium	
15	WXM Venture	Medium	
16	Hansen PLC	Low	
17	Kish Consulting	Low	→ Kish Consulting
18	Rubin Inc.	Low	
19	Sheatsley Co.	Low	
20	Steinberg Ltd.	Low	

Systematic Selection, RS = 2

Figure 4.7 Frame population of 20 establishments sorted by group with systematic selection; selection interval = 5 and random start = 2.

Muestreo Aleatorio Estratificado

- La base es la pertenencia de cada unidad a un solo estrato
- Con el MAE se puede tratar con independencia a cada estrato. Con esto se pueden utilizar distintos métodos de muestreo para cada estrato y la estimación por separado de ciertas subpoblaciones
- Lo deseable es que reúna en cada estrato a unidades homogéneas entre sí y heterogéneas en relación a otros estratos.
 - Se trata, principalmente, de reducir la varianza de las estimaciones muestrales
 - Con esto, se reducen los errores de muestreo y se gana en precisión

Muestreo Aleatorio Estratificado

- Si todo sale bien, el muestreo es más preciso que el MAS y el «efecto de diseño» inferior a 1.
 - En los hechos, para un error muestral dado, es necesario hacer un menor número de encuestas
 - Si se conoce las varianzas por estrato, es posible aplicar fracciones de muestreo distintas, en funciones del valor de la varianza (en estratos más homogéneos, fracciones de muestreo más pequeñas que en estratos más heterogéneos)

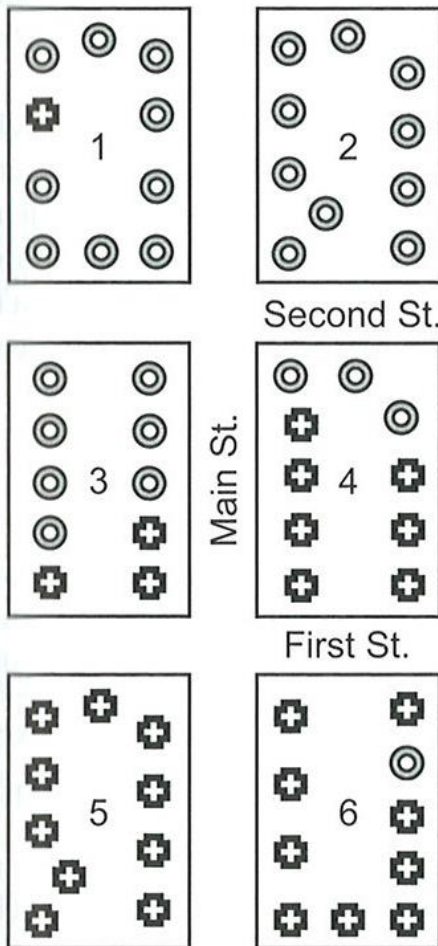
Muestreo Aleatorio Estratificado

- ¿En torno a qué variable estratificar?
 - Tiene que estar correlacionada con las variables objeto de la investigación (ej: para una encuesta sobre características de la pobreza, estratificar según ingresos)
- Aunque el equilibrio es necesario: estratos muy pequeños aportan muy poco a los beneficios de la estratificación

Muestreo Aleatorio por Conglomerados

- En este caso las unidades muestrales ya no son los individuos, sino un conjunto de individuos de los que podemos decir que forman una unidad: los conglomerados (a veces se habla de «áreas»).
 - Hospitales, liceos, cajas de productos...
- Cuando el conglomerado es pequeño, se encuesta a todas las unidades. Cuando es grande, se suele recurrir al submuestreo

Frame Population



Samples of Two Distinct Blocks
(Proportion “⊙” Households in Realization)

Block Numbers	Proportion “⊙”
1,2	$19/20 = .95^*$
2,3	$17/20 = .85^*$
1,3	$16/20 = .80^*$
2,4	$13/20 = .65$
1,4	$12/20 = .60$
2,6	$11/20 = .55$
1,6	$10/20 = .50$
2,5	$10/20 = .50$
3,4	$10/20 = .50$
1,5	$9/20 = .45$
3,6	$8/20 = .40$
3,5	$7/20 = .35$
4,6	$4/20 = .20^*$
4,5	$3/20 = .15^*$
5,6	$1/20 = .05^*$

Figure 4.4 A bird’s-eye view of a population of 30 “⊕” and 30 “⊙” households clustered into six city blocks, from which two blocks are selected.

Muestreo Aleatorio por Conglomerados Monoetápico

- En el primero de esos casos, hablamos de Muestreo Aleatorio por Conglomerados Monoetápico
 - Se procede al M. A. Simple, M. A. Sistemático o MAE, pero tomando en cuenta a los conglomerados como unidades
- Cuando no tenemos todo el marco muestral a nivel de individuos, podemos recurrir a este tipo de muestreo, para el que alcanza con tener un marco muestral de conglomerados

Muestreo Aleatorio por Conglomerados Polietápico

- Existen unidades primarias de muestreo (conglomerados), unidades secundarias...
- ¿Homogeneidad del conglomerado? (influye en el número de unidades a seleccionar)
 - Cuanto más homogeneidad, menos unidades secundarias son necesarias y se pueden tomar más conglomerados; también viceversa)
 - Se suele recomendar que las unidades primarias sean heterogéneas respecto a las características de estudio
- ¿Tamaño del conglomerado? (influye en el método de selección)
 - Si son distintos de tamaño, no deberían tener la misma probabilidad de selección

Ejemplo: investigación sobre consumos familiares

- Elección de secciones censales (unidades primarias)...
- ...dentro de las que existen rutas aleatorias para elegir hogares: una cada 100 viviendas, empezando a contar a partir de un número generado aleatoriamente (unidades secundarias)...
- ...dentro de los cuales se sortea un miembro adulto del hogar

Muestreo por cuotas

- Consiste en buscar la representatividad de la población en la última etapa. Allí las unidades no serían elegidas aleatoriamente sino cumpliendo con requisitos (habitualmente de sexo y edad, «como si fuera al azar»...)
 - Se intenta replicar en la muestra la estructura por sexo y edad de la población (que figura en el censo, por ej.) Aunque no sea un muestreo probabilístico, y por no tanto no permita (en principio) hacer estimaciones rigurosas, con medición de errores muestrales e intervalos de confianza, se suelen obtener buenos resultados

Tamaño de la muestra

- A igualdad de condiciones, mayor tamaño de la muestra implica mayor precisión de las estimaciones, pero esta relación no es lineal y directa, tal como puede pensarse.
- La precisión de las estimaciones implica un intervalo de confianza menor y, en ausencia de sesgos, menor diferencia entre las estimaciones muestrales y los parámetros poblacionales

¿Cómo determinar el tamaño de la muestra?

- **Varianza poblacional**

- Es clave la distribución del atributo en la población, más que su tamaño.

- Si hay 500 bolas negras y 500 bolas blancas en un bolillero, ¿necesitamos la misma muestra que si hubiesen 100 blancas y 900 negras?

- Dado que la primera población es más heterogénea (probabilidad de negras de 0,5 y no de 0,9), necesitamos una muestra mayor para lograr representatividad

- Si hay dos categorías, el producto de ambas probabilidades (P y $1-P$) equivale a la varianza. Cuanto más se aleje cada una de 0,5 la varianza será menor.

- Es decir: cuanto más uniforme la población, menos cantidad de encuestas son necesarias para alcanzar un nivel de precisión determinado.

- Se trata de un equilibrio entre

- Cantidad de entrevistas

- Nivel de precisión

- Varianza poblacional

¿Cómo determinar el tamaño de la muestra?

- **Tipo de muestreo**

- Comparado con el MAS...

- El MAE es más preciso (si acotamos subpoblaciones homogéneas a su interior y heterogéneas entre sí, disminuye la varianza total)
 - El MAC es menos preciso

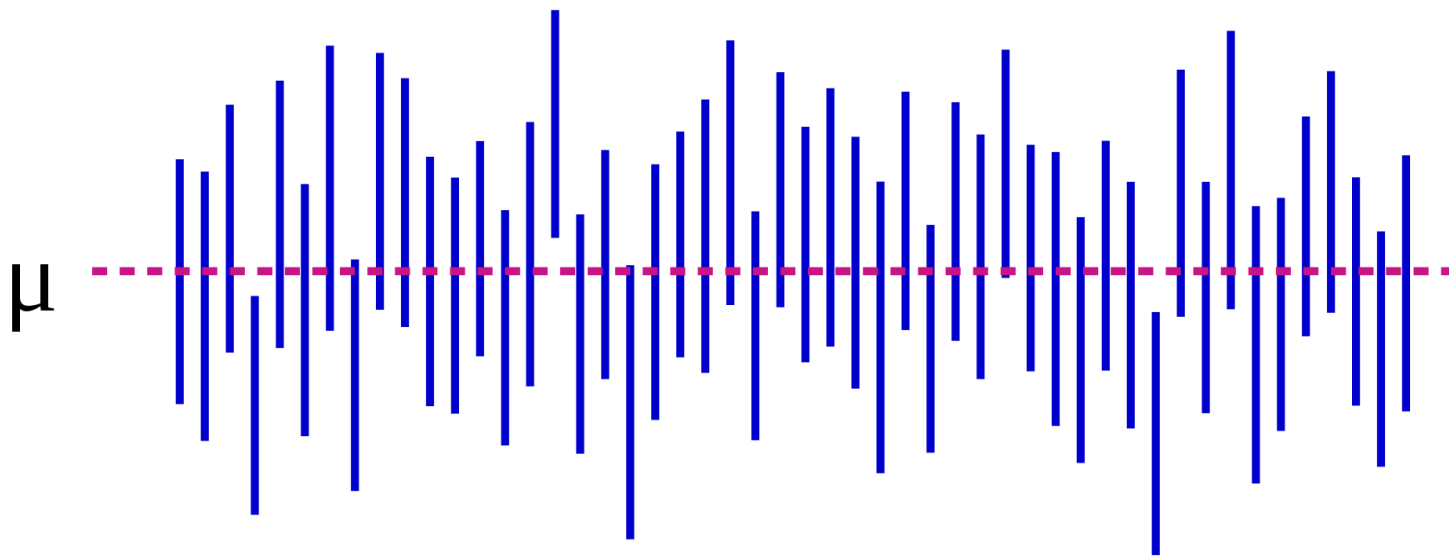
¿Cómo determinar el tamaño de la muestra?

- **Nivel de confianza**

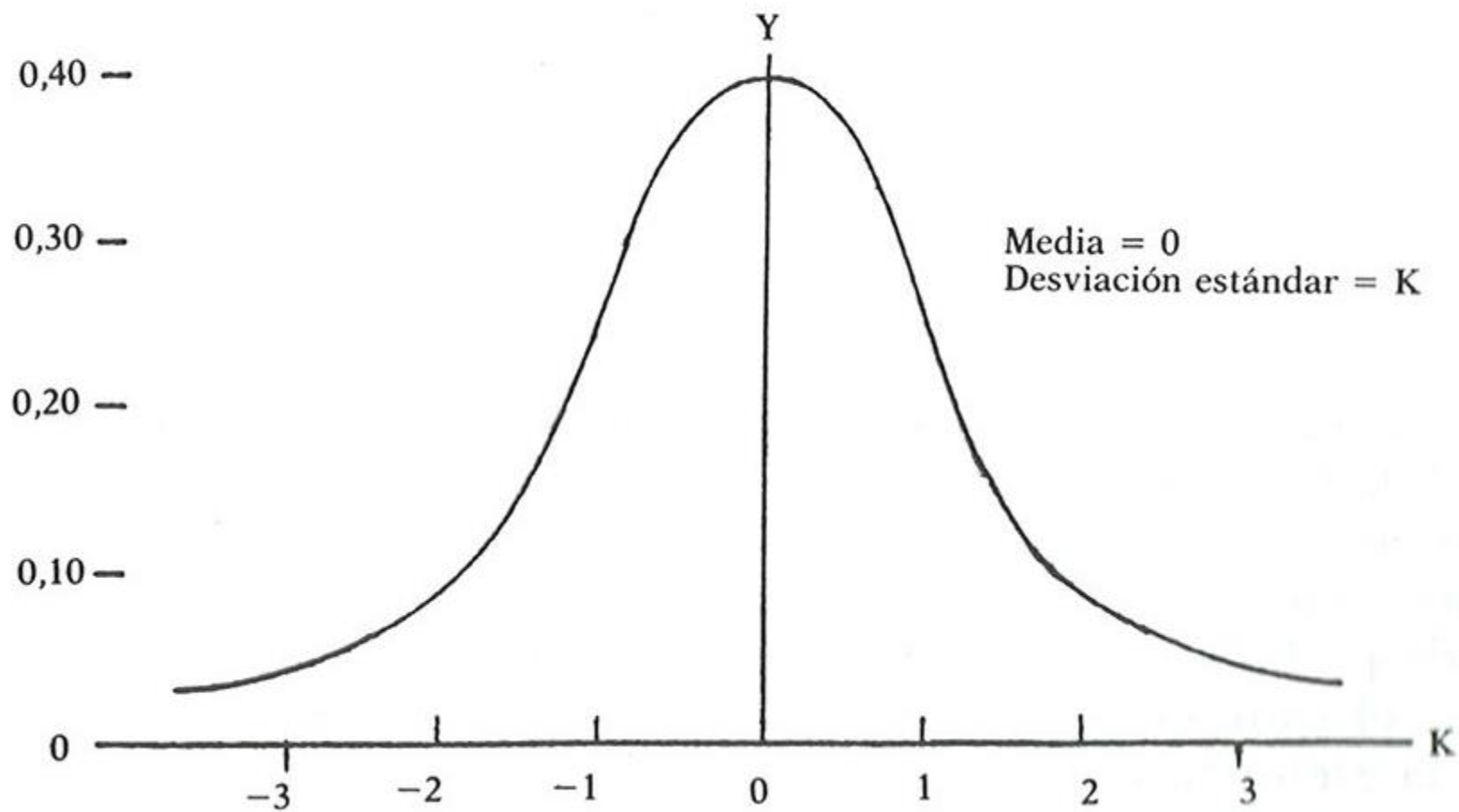
- Refiere a la *probabilidad de acertar*; a la probabilidad de que una estimación, en ausencia de sesgos, se ajuste a la realidad poblacional
- Es una acotación de la función de densidad de probabilidades, o simplemente *distribución de probabilidad* (suele ser una curva normal o *gaussiana*, sobre todo en muestra grandes)
- Las decisiones son de los investigadores, de acuerdo a la elección de uno, dos o tres desvíos estándar
- Un criterio habitual es de dos desvíos, con lo que la probabilidad de acertar en las estimaciones sería de 95,44%

Nivel de confianza

- Las líneas verticales representan 50 construcciones diferentes de intervalos de confianza para la estimación del valor μ .



CURVA NORMAL ESTÁNDAR (O UNITARIA)



¿Cómo determinar el tamaño de la muestra? Afijación de la muestra

- La afijación de la muestra es la distribución de las entrevistas entre los diferentes subconjuntos en que se divide el objeto del estudio
- Los criterios básicos son:
 - Afijación simple (cada estrato con un número igual de entrevistas)
 - Afijación proporcional (número de entrevistas de acuerdo al peso relativo de la población de cada estrato)
 - Afijación óptima (tomando en cuenta homogeneidad o heterogeneidad de la población: su varianza, o la raíz cuadrada de la varianza, llamada desviación típica)

CUADRO 3.4

APLICACIÓN DE DIFERENTES TIPOS DE AFIJACIÓN A UNA MUESTRA DE FUNCIONARIOS

<i>Estratos</i>	<i>Porcentaje funcionarios</i>	<i>Desviación típica</i>	<i>Afijación</i>		
			<i>1</i>	<i>2</i>	<i>3</i>
1. Grupo A	10	100	250	100	260
2. Grupo B	15	60	250	150	240
3. Grupo C	20	40	250	200	210
4. Grupo D	55	20	250	550	290
Total	100		1.000	1.000	1.000

NOTAS: *Afijación*: 1) simple; 2) proporcional; 3) óptima.

CUADRO 3.5

APLICACIÓN DE LA AFIJACIÓN ÓPTIMA A LA MUESTRA DE FUNCIONARIOS

<i>Estrato/Grupos</i>	<i>A. Porcentaje funcionarios</i>	<i>B. Desviación Típica</i>	<i>A × B</i>	<i>C. Porcen- taje</i>	<i>Distribución de acuerdo a C</i>
1. Grupo A.....	10	100	1.000	26	260
2. Grupo B	15	60	900	24	240
3. Grupo C.....	20	40	800	21	210
4. Grupo D	55	20	1.100	29	290
	100		3.800	100	1.000

FUENTE: Tomado de Jacinto RODRÍGUEZ OSUNA: «La muestra, teoría y aplicación» en GARCÍA FERRANDO y otros, *El análisis de la realidad social*, Madrid, Alianza Editorial, 1986, p. 279.

- Afijación óptima
 - Luego de $A \times B$, se distribuye proporcionalmente...

Afijación de la muestra y nivel de desagregación

- El error de muestreo aumenta a medida que queremos inferir conclusiones para subpoblaciones en las que el número de entrevistas es menor
 - Dado que el número de entrevistas suele ser distinto en cada subpoblación, el nivel de precisión puede ser diferente
 - ¿Qué hacer?
 - Aumentar el tamaño de la muestra total hasta que el número correspondiente a la subpoblación sea suficiente (solución costosa!)
 - Tratar a cada subpoblación como un universo y aumentar la muestra allí donde lo necesitemos (puede pasar que una ciudad necesite tantas entrevistas como el resto de un país, en la misma investigación)

Ponderación de la muestra

- Siempre que no sea proporcional la afijación, hay que ponderar la muestra: devolverle su proporcionalidad
- Ponderadores vs. expansores (elevadores)

CUADRO 3.8

CÁLCULO DEL COEFICIENTE DE PONDERACIÓN Y DE ELEVADORES

	<i>Provincias</i>		<i>Total</i>
	<i>A</i>	<i>B</i>	
1. Población de 18 años y más			
1.1. Número.....	500.000	1.000.000	1.500.000
1.2. Porcentaje.....	33,3	66,6	99,9
2. Muestra			
2.1. Elementos.....	500	500	1.000
2.2. Porcentaje sobre el total.....	50	50	100
3. Votantes del partido C en la muestra			
3.1. Número.....	300	100	400
3.2. Porcentaje sobre la muestra.....	60	20	40
4. Ponderación			
1.2 ÷ 2.2.....	0,666	1,332	
5. Elevadores			
1.1 ÷ 2.1.....	1.000	2.000	

FUENTE: Elaboración propia.

Cálculo del tamaño muestral

- Dos cálculos:
 - Para universos pequeños («poblaciones finitas»), que incluyen la «corrección por poblaciones finitas»
 - Menos de 100000 unidades
 - Para universos grandes («poblaciones finitas»), que no incluyen tal corrección
 - Más de 100000 unidades
- Los cálculos serán distintos según lo que queramos estimar (medias, totales, proporciones)

Cálculo del tamaño muestral

- La distribución de la variable objeto del estudio
- El tamaño de la población
 - Población finita (<100000 unidades)
 - Población infinita (100000 unidades o más)
- La precisión exigida
- El nivel de confianza

Para proporciones en poblaciones infinitas...

$$n = \frac{z^2 pq}{B^2}$$

- n = Tamaño de la muestra,
- z = Nivel de confianza (1,96 para el 95% de confianza, 2,56 para el 99%)
- p = Frecuencia esperada del factor a estudiar
- $q = 1 - p$
- B = Precisión o error admitido

Factor de corrección para poblaciones finitas

$$\frac{1}{n'} = \frac{1}{n} + \frac{1}{N}$$

- n' = Tamaño de la muestra necesario
- n = Tamaño de la muestra según la primera de las fórmulas
- N = Tamaño de la población

Para medias en poblaciones infinitas

$$n = \frac{z^2 s^2}{B^2}$$

- n= Tamaño de la muestra
- z= Nivel de confianza (1,96 para el 95% de confianza, 2,56 para el 99%)
- S= Desviación estándar
- B= Precisión

Tamaños muestrales según tamaño del error y distribución de la variables de interés

Prevalencia esperada	Precisión o error						
	25%	20%	10%	5%	3%	1%	0,5%
5%	3	5	19	73	292	1.825	7.300
10%	6	9	35	139	554	3.458	13.830
15%	8	13	49	196	784	4.899	19.593
20%	10	16	62	246	984	6.147	24.587
25%	12	19	73	289	1.153	7.203	28.812
30%	13	21	81	323	1.291	8.068	32.270
35%	14	22	88	350	1.399	8.740	34.959
40%	15	24	93	369	1.476	9.220	36.880
45%	16	24	96	381	1.522	9.508	38.032
50%	16	25	97	385	1.537	9.604	38.416

Tabla 1. Tamaño de muestra necesario para determinar la prevalencia en una población grande y con un nivel de confianza del 95%

Herramientas online para calcular tamaños muestrales

- http://www.netquest.com/panel_netquest/calculadora_muestras.php
- <http://www.med.unne.edu.ar/biblioteca/calculos/calculadora.htm>