Trabajo práctico 8: Muestreo y distribuciones muestrales

Apellido y nombre: ..... Comisión: ..... Fecha: .....

AL TERMINAR ESTA GUÍA

- Seleccionar una muestra aleatoria simple con una tabla de números aleatorios y una muestra sistemática.
- Repartir una muestra entre estratos con afijación igual, proporcional y de Neyman, y estimar con el estimador estratificado.
- Construir la distribución muestral de la media y comprobar que es insesgada y que su varianza es σ^2/n.
- Distinguir el error estándar del desvío estándar y usar el teorema central del límite para calcular probabilidades sobre la media y la proporción muestrales.

EJERCICIO 1 Cuarenta escuelas rurales

A mano

Un departamento tiene 40 escuelas rurales, numeradas del 01 al 40 en el listado del ministerio. Se quiere visitar 8 para relevar el estado de los edificios. La matrícula de cada escuela, en el orden del listado, es:

Escuelas 01 a 10	15	31	119	31	9	35	9	7	55	56
Escuelas 11 a 20	58	25	30	18	63	23	23	119	26	6
Escuelas 21 a 30	53	6	7	49	6	18	24	41	48	58
Escuelas 31 a 40	142	6	62	133	16	131	72	54	33	23

Para sortear se usa esta fila de una tabla de números aleatorios, que sigue de la primera línea a la segunda:

Dígitos 1 a 25	22653	28684	91341	10784	27314
Dígitos 26 a 50	70794	77994	65628	02441	17797

a) ¿Cuál es la población, la unidad de muestreo y el marco muestral? ¿Por qué no se visitan las 40?

b) Seleccioná una muestra aleatoria simple de 8 escuelas: leé los dígitos de a pares, de izquierda a derecha, quedate con los números del 01 al 40 y salteá los repetidos. ¿Qué escuelas salen?

c) Seleccioná ahora una muestra sistemática de 8 escuelas. ¿Cada cuántas escuelas se elige una? Si el arranque aleatorio, entre 1 y 5, salió 3, ¿qué escuelas salen?

d) Calculá la matrícula media de cada muestra y la de las 40 escuelas. ¿Por qué ninguna muestra da exactamente la media de la población?

EJERCICIO 2 El gasto en transporte por zonas

A mano

Una encuesta quiere estimar el gasto mensual en transporte de los 18.000 hogares de una ciudad, con una muestra de 200 hogares. Los hogares se dividen en tres zonas, y por un estudio anterior se conoce aproximadamente cuánto varía el gasto dentro de cada una (el desvío estándar, en miles de pesos):

Zona	Hogares	Desvío del gasto
Urbana	9.900	18
Periurbana	5.400	29
Rural	2.700	42

a) ¿Por qué conviene estratificar por zona en lugar de tomar una muestra aleatoria simple de toda la ciudad?

b) Repartí los 200 hogares entre las zonas con afijación igual, proporcional y óptima de Neyman.

c) Se hizo la encuesta con afijación proporcional. Las medias de gasto en cada zona dieron, en miles de pesos: urbana 46,5 (desvío 19); periurbana 65,4 (desvío 26); rural 85,4 (desvío 45). Estimá el gasto medio de la ciudad y su error estándar.

EJERCICIO 3 Todas las muestras posibles

A mano

En una cuadra viven cuatro familias. La cantidad de hijos de cada una es:

Familia	Hijos
A	6
B	5
C	3
D	1

Se toma una muestra de dos familias y se calcula el promedio de hijos.

a) Calculá la media μ y la varianza σ^2 de la población.

b) Con reposición, hay 16 muestras ordenadas posibles de dos familias (AA, AB, ..., DD). Listalas, calculá la media de cada una y armá la distribución de frecuencias de la media muestral.

c) Calculá la media y la varianza de esa distribución. Comparalas con μ y con σ^2/n.

d) Repetí sin reposición: hay 6 muestras posibles (AB, AC, AD, BC, BD, CD). ¿Sigue valiendo $E(\bar{x}) = \mu$? ¿Y la varianza?

EJERCICIO 4 La espera en la guardia

A mano

En la guardia de un hospital, el tiempo de espera hasta la primera atención tiene media 45 minutos y desvío estándar 24 minutos. La distribución es muy asimétrica a la derecha: la mayoría espera poco y algunos esperan horas.

Cada mes, la dirección del hospital revisa una muestra aleatoria de 36 pacientes y calcula el tiempo de espera promedio.

a) ¿Cuál es el error estándar de la media? ¿Qué diferencia hay entre ese número y el desvío estándar de 24 minutos?

b) ¿Se puede usar la normal para la media muestral, si los tiempos de espera son tan asimétricos?

c) ¿Cuál es la probabilidad de que el promedio de un mes supere los 55 minutos, si el funcionamiento de la guardia no cambió? ¿Y de que quede por debajo de 40?

d) ¿De cuántos pacientes tendría que ser la muestra para que el error estándar sea de 3 minutos como mucho?

e) En la misma guardia, el 30 % de los pacientes se va sin ser atendido. Si se revisa una muestra de 200 pacientes, ¿cuál es el error estándar de la proporción muestral? ¿Cuál es la probabilidad de que en la muestra se vaya más del 35 %?

EJERCICIO 5 Miles de muestras de la EPH

Con software

Se usa la base de la semana 1: «Ocupación e ingresos en Gran Catamarca», extracto de la EPH del primer trimestre de 2026 (estadistica.ar/datos/eph-catamarca-2026t1). Quedate con las personas ocupadas que declararon ingreso (ESTADO = 1 y P21 > 0).

Para este ejercicio, esas personas son la población: se conocen todas, y se puede comparar lo que dicen las muestras con el valor verdadero. En R, sample(x, n) toma una muestra sin reposición y replicate(2000, mean(sample(x, n))) repite el experimento.

a) ¿Cuántas personas tiene la población? Calculá la media, la mediana y el desvío estándar del ingreso (P21). ¿Qué forma tiene la distribución?

b) Tomá 2.000 muestras de 5 personas y otras 2.000 de 100 personas. Para cada tamaño, calculá la media de cada muestra y hacé un histograma de las 2.000 medias. ¿Qué cambia entre uno y otro?

c) Calculá el desvío estándar de las 2.000 medias para cada tamaño y comparalo con $\sigma/\sqrt{n}$. ¿Hace falta la corrección por población finita?

d) ¿Qué proporción de la población gana un millón de pesos o más? Si se toma una muestra de 100 personas, ¿cuál es el error estándar de la proporción muestral? Comprobalo con una simulación.

EJERCICIO 6 ¿Qué diseño conviene?

Para discutir

Para cada situación, decidí qué diseño de muestreo usarías (aleatorio simple, sistemático, estratificado o por conglomerados) o qué problema tiene el que se propone. Pensá en el marco muestral disponible y en el costo.

a) Se quiere estimar la proporción de estudiantes secundarios de la provincia que fuma. Hay un listado de escuelas, pero no de estudiantes.

b) Un control de calidad revisa una de cada 50 historias clínicas archivadas, en el orden del archivo.

c) Una radio hace una encuesta en su página web: «¿Está de acuerdo con el aumento del boleto?». Respondieron 3.000 personas y el 85 % dijo que no.

d) Se quiere estimar el ingreso medio de los hogares de la provincia y también dar resultados para cada departamento, incluidos los más chicos.

e) Para conocer la opinión sobre el transporte público, un encuestador entrevista a las personas que esperan en la parada de la plaza principal un martes a la mañana.
