

Estadística – Licenciatura en Nutrición - UCEL

Unidad 2: Métodos de recolección de datos

RESOLUCIÓN DE EJERCICIOS PROPUESTOS EN LAS DIAPOSITIVAS. COMISIÓN N2

Manos a la Obra N° 1

SOBREPESO EN NIÑOS DE MENOS DE 5 AÑOS

- 1) Población: todos los niños de menos de 5 años de Argentina.
- 2) Muestra: 400 niños de menos de 5 años.
- 3) Estado de sobrepeso de los niños: presente o ausente (sí o no).
- 4) El parámetro es p , la proporción de niños de menos de 5 años de Argentina que tienen sobrepeso. Su valor es $p = 0,1$.
- 5) La estadística es \hat{p} , la proporción de niños menores de 5 años en la muestra que presentan sobrepeso. Su valor es $\hat{p} = 0,12$.

¿PARÁMETRO O ESTADÍSTICA?

Este 60% es una ESTADÍSTICA y el tamaño de la muestra es $n=125$.

La población de interés es TODAS LAS CLASES DE AEROBICS DE TODOS LOS CLUBES DE SALUD DE DETROIT y el parámetro es p , LA PROPORCIÓN DE CLASES DE AEROBICS EN CLUBES DE SALUD EN LAS CUALES EL SONIDO DE LA MÚSICA SUPERA EL LÍMITE DE SEGURIDAD.

Sesgo (diapositiva 12)

Probablemente respondan a la encuesta telefónica solamente las personas que piensan que la obesidad es un problema muy serio, y que además tienen la posibilidad de pagar por la llamada telefónica (o el interés en hacer ese pago). Los resultados de esta encuesta serán sesgados, sobre todo por el sesgo en la selección de las unidades que se analizó anteriormente.

Manos a la Obra N° 2

TAMAÑO DE LA FAMILIA

- 1) Población: hogares en Estados Unidos de América.
- 2) Variable: número de personas que viven en un hogar.
- 3) Parámetro: número promedio de personas que viven en los hogares de Estados Unidos de América.

PREGUNTA SENSIBLE

- 1) En este estudio puede ocurrir un sesgo de no respuesta, es decir, que los jóvenes que no contestan la pregunta sobre su consumo de bebidas alcohólicas tengan hábitos de consumo diferentes a los de los alumnos que sí contestan esta pregunta.

- 2) Probablemente, no contesten la pregunta los alumnos que sienten que su consumo semanal de bebidas alcohólicas es excesiva, entonces se espera que el consumo promedio semanal de bebidas alcohólicas en la muestra sea inferior al consumo promedio en la población.

Manos a la Obra N° 3

SITUACIÓN NUTRICIONAL

- 1) De los 80 alumnos de la cohorte se seleccionará como muestra de tamaño $n=5$ a los individuos: 47, 53, 68, 57, 34 (el segundo 53 se elimina porque el valor está repetido).
- 2) La proporción estimada de alumnos celíacos en la cohorte es: $\hat{p} = \frac{1}{5} = 0,2$.

PARA PENSAR. PÁGINA 95

Es posible y fácil extraer una muestra simple al azar cuando contamos con un listado completo de todos los individuos (u objetos) que forman la población de interés. Si no contamos con este listado completo no podremos seleccionar la muestra usando el método del muestreo simple al azar. Además, hay restricciones de tiempo y económicas, ya que puede ser difícil poder entrevistar a cada uno de los individuos que conforman la muestra: por ejemplo, si la muestra está compuesta por 500 personas de 20 localidades diferentes deberíamos contar con entrevistadores en todas las localidades o lograr una comunicación telefónica o por correo, pero estas vías alternativas no son tan efectivas, y no hay seguridad de obtener una respuesta.

La falta de respuesta en individuos seleccionados en cualquier tipo de muestra es un problema importante, y existen numerosas técnicas para enfrentar estas situaciones, que no trataremos en esta materia.

Desventajas del MSA (diapositiva 23)

Al seleccionar una muestra simple al azar de sus pacientes en la cual por “mala suerte” no se elige a ningún paciente diabético, se va a sobreestimar el consumo de azúcar. Es decir, el consumo promedio de azúcar en la muestra de 30 pacientes será más grande que el consumo promedio de azúcar de la población de todos los pacientes de la Lic. En Nutrición, porque hay un grupo de ellos que consume menos azúcar.

Muestreo Estratificado (diapositiva 26)

En este ejemplo tenemos dos estratos:

- Estrato I, compuesto por hombres: en este estrato no hay ninguna variabilidad, todos los hombres de esta población consumen 5 porciones de carne.
- Estrato II, compuesto por mujeres: en este estrato hay variabilidad, ya que las distintas mujeres que lo conforman consumen distintas cantidades de porciones de carne: 1, 2 y 3 porciones.

Para decidir cuántas unidades seleccionar de cada estrato, usamos la información acerca de la variabilidad presente en los estratos. Como en el estrato de los hombres no hay variabilidad (o

en estratos en los que hay poca variabilidad) es suficiente con elegir un solo individuo de este estrato ya que cualquiera de ellos me da la misma información. En cambio, necesitamos más información del estrato de las mujeres para captar las diferencias de comportamiento en cuanto al consumo de carne que se observan en el mismo, por lo que elegiríamos de este estrato los restantes tres individuos para la muestra.

Conclusión: no siempre es conveniente extraer para la muestra la misma proporción de unidades de cada estrato. Por ejemplo, si tenemos 5 estratos y queremos extraer una muestra de 100 unidades, seleccionar 20 unidades de cada estrato no necesariamente será el mejor procedimiento. Si sabemos, por ejemplo, que uno de los estratos es más variable que los demás con respecto a la variable de interés, sería más conveniente seleccionar más unidades de ese estrato que de los demás: por ejemplo, 40 unidades de ese estrato y 15 de cada uno de los demás.

Aclaración importante: en general no vamos a elegir sólo un individuo de un estrato, este es un ejemplo extremo.

Manos a la Obra N° 4

LICENCIAS PARA COMEDORES ESCOLARES

Debemos seleccionar una muestra aleatoria de $n=12$ comedores escolares para que el inspector visite en el transcurso de un mes. De los $N=40$ comedores escolares que tienen licencias, 10 pertenecen al Estrato I (comedores pequeños) y 30 al Estrato II (comedores grandes). Vamos a usar muestreo estratificado para asegurarnos de tener comedores de los dos estratos en la muestra. Como el ejercicio no proporciona información sobre la variabilidad dentro de los estratos, la sugerencia es que de la muestra de $n=12$ comedores, la cantidad de unidades de los estratos I y II sean proporcionales al tamaño de los estratos:

- El estrato I de comedores pequeños está formado por el 25% de todos los comedores escolares de la ciudad ($100 \times 10 / 40$).
- El estrato II de comedores grandes está formado por el 75% de todos los comedores escolares de la ciudad ($100 \times 30 / 40$)

Usando esta información, y suponiendo que los dos estratos son igualmente variables, decidimos seleccionar para la muestra de $n=12$ comedores:

- $0,25 \times 12 = 3$ comedores del estrato I (comedores pequeños)
- $0,75 \times 12 = 9$ comedores del estrato II (comedores grandes)

Para seleccionar los 3 comedores del estrato I usamos la fila 20 y la columna 11 de la TNA. Recordando que tenemos 10 comedores en este estrato, se eligen los comedores: 2, 4, 9.

Para seleccionar los 9 comedores del estrato II usamos la fila 30 y la columna 6 de la TNA. Recordando que tenemos 30 comedores en este estrato, se eligen los comedores: 26, 17, 30, 25, 3, 22, 20, 15, 14.

PARA PENSAR. PÁGINA 100

- 1) Deben seleccionarse muestras más grandes de los estratos que tienen mayor variabilidad con respecto a la variable que se quiere estudiar.

- 2) La variabilidad dentro de los estratos debe ser menor a la variabilidad entre los estratos.
- 3) Una muestra estratificada no es una muestra simple al azar. En el muestreo estratificado se selecciona una muestra simple al azar de cada uno de los estratos.

Manos a la Obra N° 5

PARA PENSAR. PÁGINA 105

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19

Como tenemos que seleccionar una muestra sistemática de 1 en $K=4$, consideramos los dígitos en la tabla de números aleatorios de a 1, ya que solamente tenemos que seleccionar aleatoriamente el punto de partida, k , que es un número del 1 al 4.

- 1) Usando la fila 2 de la TNA, el primer elemento de la muestra es el elemento B, y los elementos restantes son F, J, N y R.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19

$k+K=2+4=6$

$k+2K=6+4=10$

$k+3K=10+4=14$

$k+4K=14+4=18$

- 2) El tamaño de la muestra es de $n=4$.
- 3) Usando la fila 4 de la TNA, el primer elemento de la muestra es el elemento D, y los elementos restantes son H, L y P.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19

$k+K=4+4=8$

$k+2K=8+4=12$

$k+3K=12+4=16$

Nos queda un individuo menos porque el individuo 20, que seguiría en la secuencia, no existe en la población.

- 4) El tamaño de la muestra es $n=3$.
- 5) El tamaño de la muestra no resulta fijo porque el K seleccionado no es proporcional al tamaño de la población, $N=19$. En el caso de este tamaño de población no se puede seleccionar un K proporcional (porque 19 es un número primo).

Veamos un ejemplo en cual todos los tamaños de muestra posibles son iguales, para cualquier arranque k . Si la población es de tamaño $N=1000$, y se decide realizar un muestreo sistemático de 1 en $K=10$, como 1000 es divisible por K (es decir, el cociente de 1000 sobre 10 tiene resto 0), para cualquier arranque aleatorio seleccionado, la muestra será de tamaño 100.

Para comprobar esta idea, piense en una población de tamaño $N=20$ y busque todas las muestras sistemáticas posibles de 1 en $K=5$. En este caso, como N es divisible por K , todas las muestras deben resultar de $n=4$.

Manos a la Obra N° 6

PARA PENSAR. PÁGINA 111

- 1) Una muestra por conglomerados no es una muestra estratificada ya que no se seleccionan individuos de cada grupo (conglomerado), sino que se selecciona una única muestra de conglomerados, y se registra la información de todas las unidades que formen parte de los conglomerados seleccionados.
- 2) La variabilidad dentro los conglomerados debe ser más grande que la variabilidad entre los conglomerados.
- 3) El criterio de la variabilidad es justamente opuesto a lo que debe ocurrir con el muestreo estratificado, en el cual la variabilidad dentro del estrato debe ser más pequeña que la variabilidad entre los estratos.

Manos a la Obra N° 7

PARA RESOLVER 2.12. Mencione el método de muestreo

- 1) En este problema se utilizó el método de muestreo por conglomerados. Las unidades son las bolsas de papas. Los conglomerados son las cajas, cada una compuesta por 20 bolsas. Se selecciona una muestra simple al azar de 5 cajas (conglomerados) y se analizan todas las bolsas (unidades) contenidas en esa muestra.
- 2) Esta es una muestra sistemática de 1 en $K=16$. Los individuos (o unidades) son los alumnos. No sabemos cuál es el tamaño de la población de alumnos, pero se decide elegir a uno de los primeros 16 alumnos de la lista, y luego ir eligiendo uno de cada 16 alumnos en el orden en que se presentan los códigos identificatorios.
Comentario: es común que, cuando se cuenta con un orden de unidades pero no se sabe con certeza el tamaño de la población, se utiliza el muestreo sistemático.
- 3) Este es un muestreo por conveniencia. Este tipo de muestreo es no probabilístico.

PARA RESOLVER 2.13. ¿Qué método de muestreo podría haber usado?

Muestra	Método de muestreo		
	Simple al azar	Estratificado	Conglomerado
Augusto, Eugenio, Gastón, Dante	X	X	X
Nora, Pablito, Augusto, Gabriel	X		
Augusto, Eugenio, Marta, Memi		X	