DEPARTAMENTO DE MATEMÁTICAS.

ESTADÍSTICA DE BACHILLERATO.

UNIDAD 6. MUESTREO E INFERENCIA.

IES LA BAHIA

1. Muestreo. Tipos de muestreo.

Como ya es sabido, no es posible estudiar cada uno de los individuos que forman la población, bien por motivos económicos o bien por la cantidad de tiempo que ello requiere. Para inferir conclusiones sobre la variable objeto de estudio de la población es necesario tomar y analizar los datos de una muestra.

Si se quiere que estas conclusiones sean fiables, es necesario que la muestra elegida sea **representativa** de la población. Al proceso de selección de una muestra se le llama **muestreo**. En particular, se llama **muestreo aleatorio** a aquel en el que cada individuo de la población tiene la misma probabilidad de ser elegido. Con este tipo de muestreo se pretende que la muestra sea representativa. El muestreo aleatorio puede ser simple o estratificado:

a) El **muestreo aleatorio simple** consiste en seleccionar al azar los **n** elementos que formarán la muestra. De ahora en adelante, se considerará que el muestreo aleatorio simple se realiza siempre **con reemplazamiento**: todo elemento seleccionado vuelve de nuevo a la población para poder volver a ser elegido. Este tipo de muestreo se puede llevar a cabo:

– Numerando los individuos estadísticos y extrayendo **n** números al azar.

– Usando una tabla de números aleatorios.

b) El **muestreo aleatorio estratificado con afijación proporcional** consiste en dividir la población en subgrupos homogéneos, llamados **estratos** (o capas) y tomar en cada uno de ellos, una muestra aleatoria simple de tamaño proporcional al tamaño del estrato.

Ejemplo: si de una población de 10 000 individuos compuesta por 1 500 jóvenes, 7 500 adultos y 1 000 ancianos, se desea tomar una muestra de tamaño 200 mediante muestreo aleatorio estratificado con afijación proporcional, habría que seleccionar al azar 30 jóvenes, 150 adultos y 20 ancianos.

2. Parámetros poblacionales y parámetros muestrales.

Parámetros poblacionales son los que definen a una población.

– la media poblacional se representa por μ

– la desviación típica poblacional se representa por σ (la varianza poblacional es σ^2)

– la proporción poblacional se representa por p

Los parámetros poblacionales son constantes. Son valores fijos de la población que se pretende estudiar.

Parámetros muestrales (o estadísticos) son los que definen a una muestra. También se les suele llamar **estimadores**.

– una media muestral se representa por $\overline{x}$

– una proporción muestral se representa por $\hat{p}$

Los parámetros muestrales, a diferencia de los poblacionales, son valores variables que dependen de la muestra elegida.

Ejemplo: se considera una población formada por un total de 93 estudiantes de ESO y sus calificaciones reales obtenidas en Matemáticas en la convocatoria ordinaria de junio.

Se numeran los 93 estudiantes usando dos cifras: 00, 01, 02, 03, 04, …, 91, 92

Alumno/a	Nota	Alumno/a	Nota	Alumno/a	Nota	Alumno/a	Nota	Alumno/a	Nota
00	3	19	2	38	6	57	6	76	6
01	5	20	8	39	2	58	3	77	7
02	1	21	0	40	3	59	8	78	9
03	6	22	1	41	5	60	3	79	5
04	1	23	5	42	8	61	5	80	2
05	8	24	6	43	2	62	7	81	6
06	3	25	7	44	7	63	3	82	8
07	9	26	8	45	6	64	7	83	2
08	5	27	5	46	6	65	2	84	7
09	4	28	3	47	0	66	5	85	5
10	7	29	8	48	5	67	6	86	9
11	2	30	10	49	3	68	9	87	8
12	6	31	6	50	8	69	7	88	6
13	3	32	2	51	5	70	3	89	5
14	5	33	3	52	2	71	5	90	4
15	7	34	5	53	10	72	3	91	8
16	1	35	0	54	5	73	10	92	5
17	9	36	3	55	4	74	2		
18	5	37	7	56	6	75	9		

La media aritmética de la población es $\mu = 5{,}13$ puntos. Este dato es real, es un valor fijo.

La proporción de aprobados de la población es $p = 64{,}52\%$. Este dato es real, es un valor fijo.

Normalmente la totalidad de los datos de la población es un valor desconocido. En este ejemplo solo son 93 datos, pero en la práctica, la población objeto de estudio puede tener cientos de miles o millones de datos imposibles de obtener.

Para obtener conclusiones fiables sobre la media aritmética de las calificaciones o sobre la proporción de aprobados se necesita elegir una muestra representativa.

Se toma una muestra de 20 individuos aplicando un muestreo aleatorio simple. Escogiendo 20 números al azar, se obtienen los siguientes números de dos cifras comprendidos entre 00 y 92:

74, 26, 13, 25, 92, 86, 53, 82, 70, 41, 65, 17, 28, 55, 32, 07, 57, 18, 06, 09

Las calificaciones correspondientes a la muestra serían por tanto:

2, 8, 3, 7, 5, 9, 10, 8, 3, 5, 2, 9, 3, 4, 2, 9, 6, 5, 3, 4

La media aritmética de esta muestra es $\overline{x} = 5{,}35$ puntos
La proporción de aprobados de esta muestra es $\hat{p} = 55\%$

Conclusión:

Los parámetros muestrales son estimadores y sus valores dependen de la muestra tomada. Obsérvese que la diferencia de la media muestral con la media poblacional (media real) es +0,22 y la diferencia de la proporción muestral con la proporción poblacional (proporción real) es –9,52%. Por esta razón, hay que desconfiar de las conclusiones obtenidas a partir de una muestra. Además, en la práctica no se puede comparar los resultados de la muestra con los reales (si se conocen los valores reales, es absurdo hacer el estudio estadístico).

Ahora bien, la Inferencia Estadística como campo de conocimiento, ha demostrado que los parámetros muestrales siguen leyes previsibles. Esto permite realizar inferencias fiables a partir de los mismos e incluso medir el riesgo que se asume al realizarlas.

3. Distribución muestral de medias.

En este apartado se considerará en todo momento que la variable aleatoria X de una población de tamaño N sigue una ley Normal $N(\mu;\sigma)$

Si de esta población se extraen todas las muestras posibles de tamaño **n**, cada una de ellas tendrá una media: $\bar{x}_1$ para la primera muestra, $\bar{x}_2$ para la segunda muestra, ... Como se hacen con reemplazamiento, en total hay N^n muestras.

Se considera la variable aleatoria

$\bar{X}$ **= medias de las muestras de tamaño n** , que asigna a cada muestra su media.

La distribución de probabilidad asociada a $\bar{X}$ se llama **distribución muestral de medias** y según el Teorema Central del Límite, sigue también una ley normal $N\left(\mu;\frac{\sigma}{\sqrt{n}}\right)$. Por lo tanto:

1. La media $\mu_{\bar{x}}$ de las medias muestrales coincide con la media poblacional μ

2. La desviación típica $\sigma_{\bar{x}}$ de las medias muestrales se obtiene dividiendo la desviación típica poblacional σ entre $\sqrt{n}$

Se va a comprobar la afirmación anterior con el siguiente ejemplo, en el que se toman números pequeños para agilizar los cálculos.

Ejemplo: una población formada por 4 alumnos de ESO [N = 4] y donde la variable aleatoria es X = calificaciones de Matemáticas en junio, que toma valores 2, 4, 8, 10

Se puede comprobar que sigue una ley Normal $N(6;\sqrt{10})$. [$\mu = 6$, $\sigma = \sqrt{10}$]

Éstas son todas las muestras posibles de tamaño n = 2. En total, son $4^2 = 16$ muestras.

Muestra	Media	Muestra	Media	Muestra	Media	Muestra	Media
{ 2, 2 }	2	{ 4, 2 }	3	{ 8, 2 }	5	{ 10, 2 }	6
{ 2, 4 }	3	{ 4, 4 }	4	{ 8, 4 }	6	{ 10, 4 }	7
{ 2, 8 }	5	{ 4, 8 }	6	{ 8, 8 }	8	{ 10, 8 }	9
{ 2, 10 }	6	{ 4, 10 }	7	{ 8, 10 }	9	{ 10, 10 }	10

La distribución muestral de medias queda reflejada en la siguiente tabla:

$\bar{X}_i$	2	3	4	5	6	7	8	9	10
p_i	1/16	2/16	1/16	2/16	4/16	2/16	1/16	2/16	1/16

Si se calculan $\mu_{\bar{x}}$ y $\sigma_{\bar{x}}$, obsérvese que: $\mu_{\bar{x}} = \mu \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

$$\mu_{\bar{x}} = 2/16 + 6/16 + 4/16 + 10/16 + 24/16 + 14/16 + 8/16 + 18/16 + 10/16 = 96/16 = 6 = \mu$$

$$\sigma_{\bar{x}} = \sqrt{(4/16 + 18/16 + 16/16 + 50/16 + 144/16 + 98/16 + 64/16 + 162/16 + 100/16) - 36} =$$

$$= \sqrt{(656/16) - 36} = \sqrt{41 - 36} = \sqrt{5} = \frac{\sqrt{10}}{\sqrt{2}} = \frac{\sigma}{\sqrt{n}}$$

Observación: cuanto mayor sea el tamaño de las muestras, menor es la desviación típica $\sigma_{\bar{x}}$, luego más alta y estrecha es la curva de Gauss y por lo tanto, menos se parecen las distribuciones poblacional y muestral.

Aplicación al cálculo de probabilidades de la distribución muestral de medias

En el siguiente ejemplo, se pone de manifiesto la diferencia de resultados al analizar los individuos uno a uno (apartado a) frente al análisis de muestras (apartado b).

Ejemplo: en el último año, el peso de los recién nacidos en una maternidad se ha distribuido según una ley normal de media $\mu = 3\,100$ gramos y desviación típica $\sigma = 150$ gramos.

a) Calcular la probabilidad de que un recién nacido pese más de 3 130 gramos.

Si se llama X = peso de los recién nacidos, X sigue una ley Normal N(3100; 150).

Por lo tanto, $P(X > 3130) = P\left(Z > \frac{3130 - 3100}{150}\right) = P(Z > 0{,}2) = 0{,}4207$

Esto significa que un 42,07% de los bebés pesa más de 3 130 gramos.

b) Calcular la probabilidad de que la media de una muestra de tamaño 100 sea mayor que 3 130 gramos.

Si se llama $\overline{X}$ = medias de las muestras de tamaño 100, $\overline{X}$ sigue una ley Normal N(3100; 15) ya que $N\left(3100; \frac{150}{\sqrt{100}}\right) = N(3\,100,\ 15)$

Por lo tanto, $P(\overline{X} > 3130) = P\left(Z > \frac{3130 - 3100}{15}\right) = P(Z > 2) = 0{,}0228$

Esto significa que sólo el 2,28% de las muestras de tamaño 100 tiene una media mayor que 3 130 gramos.

4. Estimación por un intervalo de confianza de la media poblacional.

En este apartado se considerará en todo momento que la variable aleatoria X de una población de tamaño N sigue una ley Normal $N(\mu;\sigma)$, donde σ es **conocida** y μ es **desconocida**.

Objetivo: estimar un valor fiable para la media μ a partir de la media $\overline{x}$ de una muestra de tamaño **n** y conociendo el grado de fiabilidad.

Se llama **nivel de significación** (o **de riesgo**) a la medida del **riesgo** que se asume al calcular el intervalo de confianza. Se representa por α.

Se llama **nivel de confianza** al número $N_c = 100\,(1-\alpha)$ (expresado en porcentaje)

A todo nivel de confianza le corresponde un número real llamado **valor crítico**, que se representa por $Z_{\alpha/2}$ y cumple la siguiente propiedad: $P\left(Z < Z_{\alpha/2}\right) = \dfrac{1 + \frac{N_c}{100}}{2}$

También puede utilizarse equivalentemente $P\left(Z < Z_{\alpha/2}\right) = 1 - \dfrac{\alpha}{2}$

Se llama **error máximo admisible** al radio del intervalo de confianza.

Se representa por E y se calcula así: $E = Z_{\alpha/2} \cdot \sigma_{\overline{x}} = Z_{\alpha/2} \cdot \dfrac{\sigma}{\sqrt{n}}$

Se llama **intervalo de confianza** para la media μ al intervalo en el que se estima que se encuentra la media μ con un cierto nivel de confianza fijado de antemano.

Se representa por $IC(\mu)$ y su expresión es $IC(\mu) = (\overline{x} - E,\ \overline{x} + E)$

La amplitud de $IC(\mu)$ es el doble del error máximo admisible, es decir, 2E.

Esto significa que, para todas las muestras posibles formada por **n** individuos, en el $N_c\%$ de los casos, el intervalo de confianza correspondiente incluiría el verdadero valor de la media μ

Ejemplo: una muestra aleatoria de 100 estudiantes que se presentan a la Prueba de Acceso a la Universidad revela que la media de edad es 18,1 años. Hallar un intervalo de confianza del 90% para la edad media de todos los estudiantes que se presentan a las pruebas, sabiendo que la desviación típica de la población es 0,4.

Datos: $\overline{x} = 18,1$ $\quad n = 100$ $\quad N_c = 90\%$ $\quad \sigma = 0,4$

Se calcula $Z_{\alpha/2}$ buscando en la tabla de la N(0; 1) un valor tal que $P\left(Z < Z_{\alpha/2}\right) = \dfrac{1 + \frac{90}{100}}{2} = 0,95$

lo que determina $Z_{\alpha/2} = 1,645$

Se calcula $E = 1,645 \cdot \dfrac{0,4}{\sqrt{100}} = 1,645 \cdot 0,04 = 0,0658$

Por lo tanto, $IC(\mu)$ = (18.1–0.0658, 18.1+0.0658) = (18.0342, 18.1658)

Es decir, para todas las muestras posibles de 100 individuos y con un 90% de fiabilidad, el verdadero valor de la media μ se encuentra comprendido entre 18,0342 y 18,1658 años.

Tamaño mínimo de la muestra

Si en lugar de conocer el tamaño de la muestra, se conoce el error máximo admisible y se desea determinar el tamaño muestral mínimo necesario para acotar dicho error, basta con despejar **n** en la expresión anterior, obteniéndose

$$n = \left(\frac{Z_{\alpha/2} \cdot \sigma}{E}\right)^2$$

Ejemplo: se desea realizar una investigación para estimar el peso medio de los recién nacidos en cierto hospital. Se admite un error máximo de 50 gramos con un 95% de confianza. Si por estudios anteriores se sabe que la desviación típica del peso medio de los recién nacidos es de 400 gramos, ¿qué tamaño mínimo de muestra se necesita en la investigación?

Datos: $E = 50$ $\quad N_c = 95\%$ $\quad \sigma = 400$

Se calcula $Z_{\alpha/2}$ buscando en la tabla de la N(0; 1) un valor tal que $P\left(Z < Z_{\alpha/2}\right) = \dfrac{1 + \frac{95}{100}}{2} = 0,975$

lo que determina $Z_{\alpha/2} = 1,96$

Luego el tamaño mínimo de la muestra es $n = \left(\dfrac{1,96 \cdot 400}{50}\right)^2 = 245,86 \approx 246$ recién nacidos.

5. Distribución muestral de proporciones.

En este apartado se considerará en todo momento que la variable aleatoria de una población de tamaño N solo puede tomar dos posibles valores llamados **éxito** y **fracaso**.

Se llama **proporción poblacional** y se representa por **p** a la proporción de éxito que presenta la variable aleatoria en dicha población.

La proporción de fracaso se representa por **q** , siendo **q = 1–p**

Si de esta población se extraen todas las muestras posibles de tamaño n (n > 30), en cada una de ellas habrá una proporción de éxito: $\hat{p}_1$ en la primera muestra, $\hat{p}_2$ en la segunda muestra, … Como se hacen con reemplazamiento, en total hay N^n muestras.

Se considera la variable aleatoria $\hat{P}$ = **proporciones de las muestras de tamaño n**, que asigna a cada muestra su proporción.

La distribución de probabilidad asociada a $\hat{P}$ se llama **distribución muestral de proporciones**.

Dado que para n suficientemente grande, se puede aproximar la v.a. binomial por una v.a. normal $N(n \cdot p; \sqrt{n\,p\,q})$, y según el Teorema Central del Límite, la distribución muestral de proporciones sigue también una distribución normal $N\left(p; \sqrt{\frac{p \cdot q}{n}}\right)$. Por lo tanto:

1. La media $\mu_{\hat{P}}$ de las proporciones muestrales coincide con la proporción poblacional p

2. La desviación típica $\sigma_{\hat{P}}$ de las proporciones muestrales es $\sqrt{\frac{p \cdot q}{n}}$

Se va a comprobar la afirmación anterior con el siguiente ejemplo, en el que se toman números pequeños para agilizar los cálculos.

<u>Ejemplo</u>: una población formada por 4 alumnos de ESO [N = 4] y donde la variable aleatoria es X = calificación aprobada o suspensa de Inglés en junio, que toma valores 3, 6, 7, 9.

Si se llama éxito al aprobado y fracaso al suspenso, entonces la proporción poblacional es

$$p = \frac{3}{4} = 0{,}75$$

Si se toman todas las muestras posibles de tamaño n = 2, se obtienen $4^2 = 16$ muestras.

Muestra	Proporción	Muestra	Proporción	Muestra	Proporción	Muestra	Proporción
{ 3, 3 }	0	{ 6, 3 }	0,5	{ 7, 3 }	0,5	{ 9, 3 }	0,5
{ 3, 6 }	0,5	{ 6, 6 }	1	{ 7, 6 }	1	{ 9, 6 }	1
{ 3, 7 }	0,5	{ 6, 7 }	1	{ 7, 7 }	1	{ 9, 7 }	1
{ 3, 9 }	0,5	{ 6, 9 }	1	{ 7, 9 }	1	{ 9, 9 }	1

La distribución muestral de proporciones queda reflejada en la siguiente tabla:

$\hat{P}_i$	0	0,5	1
p_i	1/16	6/16	9/16

Si se calculan $\mu_{\hat{P}}$ y $\sigma_{\hat{P}}$, obsérvese que: $\mu_{\hat{P}} = p$ $\sigma_{\hat{P}} = \sqrt{\dfrac{p \cdot q}{n}}$

$$\mu_{\hat{P}} = 0 \cdot (1/16) + 0{,}5 \cdot (6/16) + 1 \cdot (9/16) = 12/16 = 0{,}75 = p$$

$$\sigma_{\hat{P}} = \sqrt{(0^2 \cdot (1/16) + 0{,}5^2 \cdot 6/16 + 1^2 \cdot (9/16) - 0{,}75^2} = \sqrt{10{,}5/16 - 9/16} = \sqrt{0{,}09375} =$$
$$= \sqrt{\frac{0{,}75 \cdot 0{,}25}{2}} = \sqrt{\frac{p \cdot q}{n}}$$

Aplicación al cálculo de probabilidades de la distribución muestral de proporciones

Ejemplo: una máquina fabrica piezas de precisión. En su producción habitual fabrica un 3% de piezas defectuosas. Un cliente recibe una caja de 500 piezas procedentes de la fábrica.

La población está formada por el conjunto de todas las piezas producidas en la fábrica.

La v. a. sería el estado de cada pieza fabricada, que puede tomar dos valores: defectuosa (éxito) o sin defecto (fracaso). Por lo tanto, $p = 0{,}03$ $q = 0{,}97$

Obsérvese que la caja de 500 piezas es una muestra de tamaño n = 500. Como n = 500 es suficientemente grande, la v. a. $\hat{P}$ = proporciones de las muestras de tamaño 500, sigue una distribución normal $N\left(p; \sqrt{\dfrac{p \cdot q}{n}}\right) = N(0{,}03;\ 0{,}0076)$

a) ¿Cuál es la probabilidad de que encuentre más del 5% de piezas defectuosas en la caja?

$$P(\hat{P} > 0{,}05) = P\left(Z > \frac{0{,}05 - 0{,}03}{0{,}0076}\right) = P(Z > 2{,}63) = 0{,}0043$$

Esto significa que un 0,43% de las muestras de tamaño 500 tiene más del 5% de piezas defectuosas (más de 25 piezas defectuosas).

b) ¿Cuál es la probabilidad de que encuentre menos de un 1% de piezas defectuosas en la caja?

$$P(\hat{P} < 0{,}01) = P\left(Z < \frac{0{,}01 - 0{,}03}{0{,}0076}\right) = P(Z < -2{,}63) = 0{,}0043$$

Esto significa que un 0,43% de las muestras de tamaño 500 tiene menos del 1% de piezas defectuosas (menos de 5 piezas defectuosas).

6. Estimación por un intervalo de confianza de la proporción poblacional.

En este apartado se considerará en todo momento que la variable aleatoria de una población de tamaño N solo puede tomar dos posibles valores llamados **éxito** y **fracaso**.

Se llama **proporción poblacional** y se representa por **p** a la proporción de éxito que presenta la v. a. en dicha población. La proporción de fracaso se representa por **q = 1–p**

Objetivo: suponiendo que la proporción poblacional **p** es desconocida, estimar un valor fiable para **p** a partir de la proporción $\hat{p}$ de una muestra de tamaño n (n > 30) y conociendo el grado de fiabilidad.

Se llama **nivel de significación** (o **de riesgo**) a la medida del **riesgo** que se asume al calcular el intervalo de confianza. Se representa por α.

Se llama **nivel de confianza** al número $N_c = 100\,(1-\alpha)$ (expresado en porcentaje)

A todo nivel de confianza le corresponde un número real llamado **valor crítico**, que se representa por $Z_{\alpha/2}$ y cumple la siguiente propiedad: $P\left(Z < Z_{\alpha/2}\right) = \dfrac{1+\frac{N_c}{100}}{2}$

También puede utilizarse equivalentemente $P\left(Z < Z_{\alpha/2}\right) = 1 - \dfrac{\alpha}{2}$

Se llama **error máximo admisible** al radio del intervalo de confianza.

Se representa por E y se calcula así: $E = Z_{\alpha/2} \cdot \sqrt{\dfrac{\hat{p}\cdot\hat{q}}{n}}$

Se llama **intervalo de confianza** para la proporción **p** al intervalo en el que se estima que se encuentra la proporción **p** con un cierto nivel de confianza fijado de antemano.

Se representa por IC(p) y su expresión es $IC(p) = (\hat{p} - E,\ \hat{p} + E)$

Esto significa, que si se obtuvieran un gran número de muestras de tamaño n, en el N_c% de los casos, el intervalo de confianza correspondiente incluiría el valor real de la proporción **p**

Ejemplo: para estimar la proporción de estudiantes de una universidad que está a favor de la energía nuclear, se entrevistó aleatoriamente a 500 estudiantes. La muestra dio un resultado de un 58% a favor y un 42% en contra. Calcular el intervalo de confianza, con un 95% de fiabilidad, en el cual se encuentra la proporción real de población que se encuentra a favor.

Datos: $\hat{p} = 0{,}58$ $\quad \hat{q} = 0{,}42$ $\quad n = 500$ $\quad N_C = 95\%$

Se calcula $Z_{\alpha/2}$ buscando en la tabla de la N(0; 1) un valor tal que $P\left(Z < Z_{\alpha/2}\right) = \dfrac{1+\frac{95}{100}}{2} = 0{,}975$

lo que determina $Z_{\alpha/2} = 1{,}96$

Se calcula $E = 1{,}96 \cdot \sqrt{\dfrac{0{,}58 \cdot 0{,}42}{500}} = 0{,}043$

Luego IC(p) = (0.58–0.043, 0.58+0.043) = (0.537, 0.623)

De este resultado se puede afirmar que, con una fiabilidad del 95%, la proporción real de población que está a favor se encuentra entre un 53,7% y un 62,3%

Tamaño mínimo de la muestra

Si en lugar de conocer el tamaño de la muestra, se conoce el error máximo admisible y se desea averiguar el tamaño muestral mínimo necesario para acotar dicho error, basta con despejar **n** en la expresión anterior, obteniéndose

$$n = \left(Z_{\alpha/2}\right)^2 \cdot \frac{\hat{p} \cdot \hat{q}}{E^2}$$

<u>Ejemplo</u>: se quiere estimar el porcentaje de conexión a Internet del que disponen los hogares españoles, con un nivel de confianza del 99,73% y con un error máximo de 0,05. Se dispone de la proporción del último estudio que fue de un 32% los que sí tenían conexión y un 68% los que no tenían. ¿Qué tamaño mínimo de muestra se necesita en el estudio?

Datos: $E = 0{,}05$ $\quad N_c\% = 99{,}73\%$ $\quad \hat{p} = 0{,}32$ $\quad \hat{q} = 0{,}68$

Se calcula $Z_{\alpha/2}$ buscando en la tabla de la N(0; 1) un valor tal que $P\left(Z < Z_{\alpha/2}\right) = \dfrac{1 + \frac{99{,}73}{100}}{2} = 0{,}99865$ lo que determina $Z_{\alpha/2} = 2{,}99$

Luego el tamaño mínimo de la muestra es $n = 2{,}99^2 \cdot \dfrac{0{,}32 \cdot 0{,}68}{0{,}05^2} = 778{,}15 \approx 779$ hogares.