

MATEMATICAS (MAS II)
Bachillerato
Inferencia. Teoría de muestras(I).



Departamento de Matemáticas

Ies Dionisio Aguado

1. TEORÍA DE MUESTRAS

1.1. Población y muestra

Entendemos por población un conjunto de elementos que poseen una característica o propiedad común, y que constituyen la totalidad de los individuos de interés para nuestro estudio.

En particular, nos interesa obtener información acerca de algún valor que caracteriza a la población, como una media, una varianza, una mediana, etc.

A estos valores que se refieren a la totalidad de la población se les denomina **parámetros poblacionales** y en su notación es común utilizar el alfabeto griego:

μ (media poblacional),

σ^2 (varianza poblacional),

σ (desviación típica poblacional), etc.

Como las poblaciones en las que se pretende estudiar una determinada variable aleatoria son grandes, es muy caro o imposible estudiar a todos sus individuos; lo que se hace, es estudiar una muestra (una parte) de la población:

DEFINICIÓN: Una **muestra** es cualquier subconjunto representativo de la población sobre el que se realizan estudios para obtener conclusiones acerca de las características de la población.

Cualquier valor obtenido a partir de los datos de la muestra se denomina **estadístico muestral**.

Ejemplos de **estadísticos muestrales** son:

\bar{x} (media muestral),

s^2 (varianza muestral),

s (desviación típica muestral), etc.

Si por ejemplo decimos que la duración media de las bombillas que fabrica una empresa es de 1600 horas, nos referimos a la población y la media la designamos por $\mu = 1600$. Sin embargo si hallamos la duración media de una muestra de 20 bombillas y obtenemos 1580 horas, nos referimos a la muestra y la media la designamos por $\bar{x}=1580$. Podríamos a continuación hallar la duración media de otra muestra de 30 bombillas y obtener 1610 horas, nos referimos a la muestra y la media la designamos por $\bar{x}=1610$.

- La **media muestral** (de las muestras) es una **variable aleatoria** mientras que la media poblacional es una constante.
- La media μ de la población es un parámetro y es constante.

- La media \bar{x} de la muestra es un estadístico y además es una variable aleatoria.

La teoría del muestreo tiene por objetivo, el estudio de las relaciones existentes entre la distribución de un carácter en dicha población y las distribuciones de dicho carácter en todas sus muestras.

Las ventajas de estudiar una población a partir de sus muestras son principalmente:

- **Coste reducido:** Si los datos que buscamos los podemos obtener a partir de una pequeña parte del total de la población, los gastos de recogida y tratamiento de los datos serán menores. Por ejemplo, cuando se realizan encuestas previas a un referéndum, es más barato preguntar a 4.000 personas su intención de voto, que a 30.000.000;
- **Mayor rapidez:** Estamos acostumbrados a ver cómo con los resultados del escrutinio de las primeras mesas electorales, se obtiene una aproximación bastante buena del resultado final de unas elecciones, muchas horas antes de que el recuento final de votos haya finalizado;
- **Más posibilidades:** Para hacer cierto tipo de estudios, por ejemplo el de duración de cierto tipo de bombillas, no es posible en la práctica destruirlas todas para conocer su vida media, ya que no quedaría nada que vender. Es mejor destruir sólo una pequeña parte de ellas y sacar conclusiones sobre las demás.

De este modo se ve que al hacer estadística inferencial debemos enfrentarnos con dos problemas:

- Elección de la muestra (**muestreo**).
- **Extrapolación de las conclusiones** obtenidas sobre la muestra, al resto de la población (**inferencia**).
 - Es decir, el principal objetivo de la mayoría de los estudios, análisis o investigaciones, es hacer generalizaciones “acertadas” con base en muestras de poblaciones de las que se derivan tales muestras. Obsérvese la palabra “acertadas” porque no es fácil responder cuándo y en qué condiciones las muestras permiten tales generalizaciones.

Por ejemplo si queremos calcular la cantidad de dinero que se gasta una persona en vacaciones, ¿tomaríamos como muestra lo que gastan los viajeros que lo hacen en primera clase? Es obvio que no, pero saber a que tipo de personas debemos incluir en nuestra muestra no es algo intuitivo ni evidente.

Sobre las muestras hay dos aspectos que resultan fundamentales:

- El tamaño (n^o de elementos de la muestra)
- La forma en que se realiza la selección de los individuos que la forman.

La elección de la muestra influirá de manera determinante en los resultados, por lo que hay que evitar muestras “sesgadas” (parciales y subjetivas) que produzcan errores incontrolables (aparte de los errores propios de sustituir el estudio de una población por el de una muestra).

El tipo de muestreo más importante es el muestreo aleatorio, en el que todos los elementos de la población tienen la misma probabilidad de ser extraídos; aunque dependiendo del problema y con el objetivo de reducir los costes o aumentar la precisión, otros tipos de muestreo pueden ser considerados como veremos a continuación.

2. TIPOS DE MUESTREO

2.1. Muestreo Aleatorio Simple

Es la técnica de muestreo en la que todos los elementos que forman el universo y que, por lo tanto, están descritos en el marco muestral, tienen idéntica probabilidad de ser seleccionados para la muestra.

Consiste en seleccionar n elementos **sin reemplazamiento** de entre los N que componen la población, de tal modo que todas las muestras tengan la misma probabilidad de ser elegidas.

Es algo así como hacer un sorteo justo entre los individuos del universo: asignamos a cada persona un boleto con un número correlativo, introducimos los números en una urna y empezamos a extraer al azar boletos. Todos los individuos que tengan un número extraído de la urna formarían la muestra. Obviamente, en la práctica estos métodos pueden automatizarse mediante el uso de ordenadores.

En la práctica, se enumeran los individuos y se sorteán cuáles de ellos se elegirán. Si los individuos son, por ejemplo, tornillos que se encuentran en una caja, se eligen al azar por simple extracción.

2.2. Muestreo Aleatorio Sistemático

Se numeran los individuos y, a partir de uno de ellos elegido al azar, se toman los siguientes mediante “saltos” numéricos iguales. El salto se llama coeficiente de elevación, k , y se obtiene mediante el cociente entero entre el número de individuos de la población, N y el número de individuos de la muestra: $n \mapsto k = \frac{N}{n}$. El primer elemento, llamado origen, se elige al azar entre los números $1, 2, 3, \dots, k$. Esta forma de muestreo sólo es válida si el criterio por el que se han numerado los individuos de la población no tiene nada que ver con la característica que se quiere estudiar con la muestra..

$$k = \frac{N \text{ de elementos de la población}}{\text{Tamaño de la muestra}} = \frac{N}{n}$$

Por ejemplo:

Si tenemos una población formada por 100 elementos y queremos extraer una muestra de 25 elementos, en primer lugar debemos establecer el intervalo

de selección que será igual a $\frac{100}{25} = 4$. A continuación elegimos el elemento de arranque, tomando aleatoriamente un número entre el 1 y el 4, (por ejemplo el 2) y a partir de él obtenemos los restantes elementos de la muestra:

$$2, 6, 10, 14, \dots, 98$$

2.3. Muestreo Aleatorio Estratificado

Si conocemos que la población puede dividirse en partes o estratos, en relación con variables que pueden ser de interés en nuestro estudio, de modo que en cada uno de los estratos los elementos posean una gran homogeneidad respecto al carácter que se estudia (sexo, grupos de edad, nivel de estudios, ...), se puede aumentar la precisión si muestreamos los estratos por separado.

La forma de repartir los elementos de la muestra, determinando cuantos deben corresponder a cada estrato se denomina afijación, y puede ser de varios tipos:

Afijación uniforme : Si se toma el mismo número de elementos en cada estrato.

$$n_1 = n_2 = \dots = n_k = \frac{n}{k}$$

Afijación Proporcional : si el número de elementos que se toma en cada estrato es proporcional al tamaño del estrato. Se utiliza cuando las varianzas de los estratos no difieren mucho entre si.

$$\frac{n_1}{N_1} = \frac{n_2}{N_2} = \dots = \frac{n_k}{N_k} = \frac{n}{N}$$

Ejemplo:

Por ejemplo, queremos estudiar el % de la población que fuma en determinado país y pensamos que la edad puede ser un buen criterio para estratificar (es decir, pensamos que existen diferencias importantes en el hábito de fumar dependiendo de la edad).

Definimos 3 estratos: menores de 20 años, de 20 a 44 años y mayores de 44 años. Es de esperar que al dividir toda la población mexicana en estos 3 estratos no resulten grupos de igual tamaño. Efectivamente, si miramos datos oficiales, obtenemos:

- Estrato 1 - Población Mexicana menor de 19 años: 42,4 millones (41,0 %)
- Estrato 2 - Población Mexicana de 20 a 44 años: 37,6 millones (36,3 %)
- Estrato 3 - Población Mexicana mayor de 44 años: 23,5 millones (22,7 %)

Si usamos muestreo estratificado proporcionado, la muestra deberá tener estratos que guarden las mismas proporciones observadas en la población. Si en este ejemplo queremos crear una muestra de 1.000 individuos, los estratos tendrán que tener un tamaño como sigue:

Estrato	Población	Proporción	Muestra
1	42,4Millones	41,0 %	410
2	37,6M	36,3 %	363
3	23,5M	22,7 %	227

Hablaremos de una afijación uniforme cuando asignamos el mismo tamaño de muestra a todos los estratos definidos, sin importar el peso que tienen esos estratos en la población. Siguiendo con el ejemplo anterior, un muestreo estratificado uniforme definiría la siguiente muestra por estrato:

Estrato	Población	Proporción	Muestra
1	42,4Millones	41,0 %	333
2	37,6Millones	36,3 %	333
3	23,5Millones	22,7 %	333

Ejemplo 2

En una ciudad se quiere hacer un estudio para conocer qué tipo de actividades se realizan en el tiempo dedicado al ocio. Para ello van a ser encuestadas 300 personas elegidas al azar mediante muestreo aleatorio estratificado con afijación proporcional. Teniendo en cuenta que de un total de 15.000 habitantes, 7.500 son adultos, 3.000 ancianos y 4.500 niños, definir los estratos y el tamaño muestral correspondiente a cada estrato.

Solución:

Basta con hacer una simple regla de 3 para determinar el tamaño muestral de cada estrato:

$$\text{Niños: } \frac{15.000}{4500} = \frac{300}{x}$$

$$x = (4.500 \cdot 300) / 15000 = 90 \text{ niños}$$

$$\text{Adultos: } 7.500 * 300 / 15000 = 150 \text{ adultos}$$

$$\text{Ancianos: } 3.000 * 300 / 15000 = 90 \text{ ancianos}$$

2.4. Muestreo Por Conglomerados o áreas

En este tipo de muestreo, llamado **muestreo por conglomerados**, se divide la población total en un número determinado de subdivisiones relativamente pequeñas y se seleccionan al azar algunas de estas subdivisiones o conglomerados, para incluirlos en la muestra total. Si estos conglomerados coinciden con áreas geográficas, este muestreo se llama también muestreo por áreas .

Por ejemplo, supongamos que una gran empresa quiere estudiar los patrones variables de los gastos familiares de una ciudad como Granada. Al intentar elaborar los programas de gastos de una muestra de 1200 familias, nos encontramos con la dificultad de realizar un muestreo aleatorio simple, (es complicado tener una lista actualizada de todos los habitantes de una ciudad). Una manera de tomar una muestra en esta situación es dividir el área total (Granada en este caso) en áreas más pequeñas que no se solapen (Por ejemplo Distritos postales, manzanas etc...) En este caso seleccionaríamos algunas áreas al azar y todas las familias (o muestras de éstas) que residen en estos distritos postales o manzanas, constituirían la muestra definitiva.

Aunque las estimaciones basadas en el muestreo por conglomerados, por lo general no son tan fiables como las obtenidas por muestreos aleatorios simples del mismo tamaño, son más baratas. Volviendo al ejemplo anterior, es mucho más económico visitar a familias que viven en el mismo vecindario, que ir visitando a familias que viven en un área muy extensa.

En la práctica se pueden combinar el uso de varios de los métodos de muestreo que hemos analizados para un mismo estudio.

Ejemplo

Se estudia la incidencia de enfermedades cardíacas en la población rural española. Para ello se hace un censo de pueblos y se eligen varios al azar, donde se estudia a la población

Ejercicios:

1. Consideremos la población formada por 5 bolas contenidas en una urna y numeradas del 1 al 5. Obtener todas las muestras de tamaño 2 extraídas mediante muestreo aleatorio simple.
2. En un instituto de enseñanza secundaria en que se ofertan los siguientes tipos de enseñanza :
 - a) Ciclos de grado superior: 110 alumnos. Bachillerato: 162 alumnos.
Ciclos de grado medio : 210 alumnos 2º ciclo de enseñanza secundaria obligatoria: 338 alumnos.

Se pretende valorar las faltas de ortografía que cometen los alumnos del centro mediante una prueba-dictado de un texto de 20 líneas; la prueba se pasará a una muestra de 50 alumnos, para minimizar el costo en tiempo y medios.

¿Cómo se obtiene una muestra adecuada a esta población?

Solución : En esta situación parece conveniente utilizar para la extracción de la muestra el muestreo aleatorio estratificado con afijación proporcional. Dividimos la población en cuatro estratos: ciclos de grado superior, ciclos de grado medio, bachillerato y 2º ciclo de enseñanza secundaria obligatoria.

Como el número total de alumnos son 820 y la muestra debe estar formada por 50 alumnos, el cálculo del número de alumnos que se han de tomar de cada estrato es:

$$\text{Ciclos de grado superior: } \frac{820}{110} = \frac{50}{x}; x = \frac{50 \cdot 110}{820} = 7$$

$$\text{Bachillerato: } \frac{820}{162} = \frac{50}{x}; x = \frac{50 \cdot 162}{820} = 10$$

$$\text{Ciclos de grado medio: } \frac{820}{210} = \frac{50}{x}; x = \frac{50 \cdot 210}{820} = 13$$

$$\text{2º ciclo de Enseñanza Secundaria Obligatoria: } \frac{820}{338} = \frac{50}{x}; x = \frac{50 \cdot 338}{820} = 20$$

3. Los empleados de una empresa están clasificados como figura en la siguiente tabla :

CATEGORIA	A	B	C	D
Nº de empleado	500	300	350	150

- a) Para hacer una consulta sobre la modificación del horario laboral, se elige por sorteo a 40 empleados de la categoría A, 30 de la B, 50 de la C y 20 de la D. ¿Es este un modelo de muestreo aleatorio estratificado? ¿Es proporcional?

Es un muestreo aleatorio estratificado pero no proporcional ya que $\frac{40}{500} \neq \frac{30}{300} \neq \frac{50}{350} \neq \frac{20}{150}$

Si quisieramos que fuera proporcional con una muestra de 140

$$\frac{A}{500} = \frac{B}{300} = \frac{C}{350} = \frac{D}{150} = \frac{140}{500} \Rightarrow A = 54; B = 32; C = 38; D = 16$$

4. En un centro, hay 2000 alumnos, 720 en 3º de ESO, 700 en 4º de ESO, 340 en 1º de Bachillerato, y 240 en 2º de Bachillerato. Si deseamos tomar una muestra de 100 alumnos, para conocer la opinión que tiene el alumnado sobre una medida que ha tomado el Consejo Escolar, ¿cómo elegirías una muestra de 100 alumnos por muestreo aleatorio estratificado?

- a) Solución: 3º de ESO: 36 alumnos; 4º de ESO: 35 alumnos; 1º de Bachillerato: 17 alumnos; 2º de Bachillerato: 12 alumnos

MUESTREO	ALEATORIO	Simple	Con Reposición/Sin reposición
		Sistemático	
		Estratificado	
		Conglomerado	
	NO ALEATORIO		

3. Tamaño y representatividad de una muestra

Cuando se elige una muestra los dos aspectos que hay que tener en cuenta son, el tamaño y la representatividad de la muestra. Si la muestra es demasiado pequeña, aunque esté bien elegida, el resultado no será fiable.

Ejemplo:

Queremos estudiar la estatura de la población española. Para ello elegimos a una persona al azar y la medimos.

Evidentemente este resultado no es fiable. La muestra es demasiado pequeña.

Si la muestra es demasiado grande los resultados serán muy fiables, pero el gasto puede ser demasiado elevado. Incluso, en ocasiones, muestras demasiado grandes no nos proporcionan mejores resultados. Vamos a aprender

a encontrar cuál es el tamaño adecuado para que podamos afirmar que la población tiene tal característica con una probabilidad dada, grande.

Cuando una muestra tenga el tamaño adecuado, y haya sido elegida de forma aleatoria diremos que es una muestra representativa.

Si la muestra no ha sido elegida de forma aleatoria diremos que la muestra es sesgada.

Ejemplo:

Indica si es población o muestra:

En una ganadería se mejora el pienso de todas las ovejas con un determinado tipo de grano. En otra ganadería se seleccionan 100 ovejas para alimentarlas con ese tipo de grano y estudiar su eficacia.

En el primer caso, todas las ovejas, son la población. En el segundo se ha elegido una muestra.

En una serie de televisión tienen dudas sobre qué hacer con la protagonista, si que tenga un accidente o si debe casarse. Van a hacer una consulta. ¿A toda la población o seleccionado una muestra representativa?

Observa que no sabemos bien cuál sería la población, ¿los que ven esa serie? o ¿toda la población española? Si son los que ven la serie, ¿cómo los conocemos? ¿Cómo preguntar a todos? Parece más operativo preguntar a una muestra.

El estudio de la vida media de unas bombillas, ¿se puede hacer sobre toda la población?

El estudio es destructivo. Si se hiciera sobre toda la población nos quedamos sin bombillas. Es imprescindible tomar una muestra.

Disponemos del censo electoral de una población. Consta de 33.600 electores. Deseamos extraer una muestra de 300 individuos.

- a) ¿Cómo se debe realizar mediante muestreo aleatorio sistemático?
- b) ¿Cómo se debe realizar mediante muestreo aleatorio simple?

Solución

a) Coeficiente de elevación : $h = \frac{33600}{300} = 112$

Esto significa que hemos de seleccionar un individuo de cada 112. Elegimos al azar un número del 1 al 112. El segundo elemento será $n + 112$, y así sucesivamente.

b) Habrá que seleccionar un individuo al azar de entre los 33600, descartando a aquellos que salgan repetido. Este proceso habrá que repetirlo 300 veces, que es el tamaño de mi muestra.

4. DISTRIBUCIÓN DE LAS MEDIAS MUESTRALES

Supongamos que queremos estudiar la media μ de una variable aleatoria X (estatura, peso, presión sanguínea, etc..) en cierta población.

Consideramos entonces el conjunto de muestras del mismo tamaño, n , extraídas de la población y definimos la variable aleatoria X que asigna a cada muestra, M_i el valor de su media, \bar{x}_i ;

$$M_i x_i \longrightarrow \bar{x}_i$$

La variable aleatoria \bar{X} se denomina **media muestral**, y su distribución de probabilidad, **distribución muestral de las medias**.

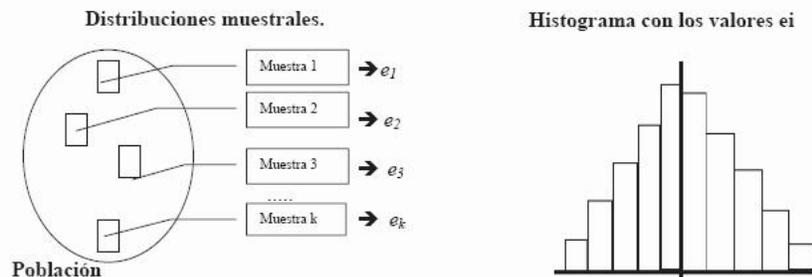
Es decir: De una población se selecciona una muestra M_i y se calcula su media \bar{x}_i y su desviación típica, s_i . Elegimos otras muestras de la misma población, y de cada una obtenemos su media y desviación típica. ¿Cómo es la distribución de esas medias? ¿Y de esas desviaciones típicas?

Los diferentes valores de las medias de las diferentes muestras dan lugar a una variable aleatoria que la vamos a representar por \bar{X} .

Con lo que la idea de **distribución muestral** es la siguiente:

Partimos de una población de tamaño N . Obtenemos k muestras (todas las posibles de tamaño n) y de cada una de ellas se calcula una medida (media, mediana, varianza, desviación típica,..) obteniendo k valores: $e_1, e_2, e_3, \dots, e_k$.

Estos valores pueden representarse mediante un histograma para observar su distribución:



Se puede observar que este histograma va adquiriendo forma de campana de Gauss, y a medida que k aumenta, **se va pareciendo cada vez más a la forma de una distribución Normal**.

Lo vemos con un ejemplo:

Elaboraremos la distribución muestral de la media de una muestra aleatoria de tamaño $n = 2$ tomada sin reemplazo de la población finita de tamaño $N = 5$, cuyos elementos son: 3,5,7,9,11.

$$\mu = \frac{3+5+7+9+11}{5} = 7; \sigma^2 = \frac{3^2+5^2+7^2+9^2+11^2}{5} - 7^2 = 8 \Rightarrow \sigma = \sqrt{8}$$

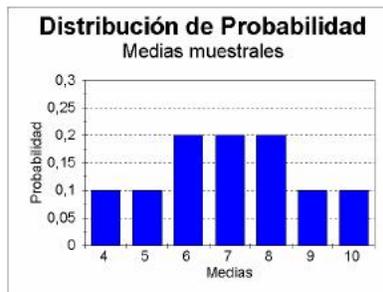
Ahora si tomamos una muestra aleatoria de tamaño $n = 2$ de esta población hay 20 posibilidades:

nº muestra	Muestras		\bar{X} (media de la muestra)
1	3	5	4
2	3	7	5
3	3	9	6
4	3	11	7
5	5	3	4
6	5	7	6
7	5	9	7
8	5	11	8
9	7	3	5
10	7	5	6
11	7	9	8
12	7	11	9
13	9	3	6
14	9	5	7
15	9	7	8
16	9	11	10
17	11	3	7
18	11	5	8
19	11	7	9
20	11	9	10

Si hacemos la función de probabilidad de la variable \bar{X} de las medias muestrales:

Media	Probabilidad
4	2/20
5	2/20
6	4/20
7	4/20
8	4/20
9	2/20
10	2/20

Cuyo histograma sería:



Si calculamos la media y la desviación típica de la distribución de las medias obtenemos que: $\mu_{\bar{x}} = 7$ y $\sigma_{\bar{x}} = \sqrt{3}$, luego la media $\mu_{\bar{x}}$ coincide con la media de la población y la desviación típica ha disminuido. Este ejemplo puede generalizarse para cualquier distribución según el siguiente teorema:

Teorema Central del Límite Si una población tiene media μ y desviación típica σ , y tomamos muestras de tamaño n ($n > 30$, ó cualquier tamaño si la población es "normal"), las medias de estas muestras siguen aproximadamente la distribución: $\bar{X} \rightarrow N(\mu, \frac{\sigma}{\sqrt{n}})$

Destacar que si la población de la que se obtienen las muestras es normal, las medias muestrales también se distribuyen según una distribución normal, independientemente del tamaño de la muestra. Al saber cómo funcionan las medias muestrales, podemos obtener probabilidades relacionadas con ellas.

Entonces :

Dada una población en la que la variable aleatoria X sigue una distribución con media μ y desviación típica σ se verifica:

- La variable aleatoria \bar{X} sigue una distribución de media μ y desviación típica $\frac{\sigma}{\sqrt{n}}$, siendo n el tamaño de las muestras, es decir:

$$\mu_{\bar{x}} = \mu \text{ y } \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

- Si la distribución de X es normal, la de \bar{X} es también normal.
- Si la distribución de X no es normal, a medida que aumenta el tamaño n de las muestras, la distribución de \bar{X} se aproxima a una distribución normal, siendo la aproximación suficientemente buena para $n \geq 30$.

Eemplo

Un vendedor sabe que el peso de los melones que vende sigue una distribución normal con peso medio $\mu = 4kg$ y desviación típica $\sigma = 250g$. ¿Qué tipo de distribución es la distribución de las medias de las muestras de 25 melones seleccionados al azar?

La distribución de la población es normal. Luego, por el resultado anterior, la distribución de las medias de las muestras de 25 melones es una distribución normal con media $\mu_{\bar{X}} = \mu = 4kg$ y desviación típica $\sigma_{\bar{X}} = \frac{250g}{\sqrt{25}} = 50g$.

Ejemplo

Se sabe que las bolsas de azúcar producidas por una máquina tiene una media de 500 g. y una desviación típica de 35 g. Dichas bolsas se empaquetan en cajas de 100 unidades.

- ¿Cómo se distribuyen las medias de los pesos de las bolsas de cada caja?
- Calcular la probabilidad de que la media de los pesos de las bolsas de un paquete sea inferior a 495 g.

Solución:

Tenemos $\mu=500, \sigma=35, n=100$;

Por el Teorema Central del Límite (**aunque la variable peso no sea normal**, tenemos $n > 30$), la variable $X =$ "media de los pesos de la muestra" se distribuye según una normal:

$$X \rightarrow N(\mu, \frac{\sigma}{\sqrt{n}}) = N(500, \frac{35}{\sqrt{100}}) = N(500, 3'5)$$

$$P(\bar{X} < 495) = \text{tipificando} = P(\frac{\bar{X}-500}{3'5} < \frac{495-500}{3'5}) = P(Z < -1'43) = P(Z < -1'43) = 1 - P(Z < 1'43) = 0,0764$$

Lo que vendría a significar que en el 7'6% de las cajas el peso medio de las bolsas es inferior a 495 g.

Ejemplo

Se supone que la estatura de los universitarios varones de una determinada ciudad sigue una distribución normal de media 176 cm y desviación típica 14 cm. Se toma una muestra al azar de 49 de estos alumnos y se calcula su estatura media.

- ¿Cuál es la probabilidad de que esta media sea menor de 174 cm?
- ¿Cuál es la probabilidad de que esté entre 174 cm y 180 cm?
- ¿Cuál es la probabilidad de que sea superior a 178 cm?

Solución:

Sea la variable aleatoria X : estatura (cm). Sabemos que $X = N(\mu, \sigma)$, con $\mu = 176$ cm y $\sigma = 14$ cm. Luego la media muestral \bar{X} sigue una distribución normal $N(\mu, \frac{\sigma}{\sqrt{n}}) = N(176, \frac{14}{\sqrt{49}}) = N(176, 2)$.

a) Tenemos que hallar la probabilidad de que la media muestral sea menor que 174, es decir, $P[\bar{X} < 174]$. Para ello, tipificamos la variable:

$$Z = \frac{\bar{X} - 176}{2} \sim N(0, 1)$$

$$P[\bar{X} < 174] = P[Z < \frac{174-176}{2}] = P[Z < -1] = P[Z > 1] = 1 - P[Z \leq 1] = 1 - 0,8413 = 0,1587 \Rightarrow 15,87\%$$

b) Debemos calcular $P[174 < X < 180]$:

$$P[174 < \bar{X} < 180] = P[\frac{174-176}{2} < Z < \frac{180-176}{2}] = P[-1 < Z < 2] = P[Z < 2] - P[Z > -1] = P[Z < 2] - (1 - P[Z < 1]) = 0,9772 - (1 - 0,8413) = 0,81851,85$$

$$c) \text{ Calculamos } P[\bar{X} > 178] = P[z > \frac{178-176}{2}] = P[Z > 1] = 1 - P[Z \leq 1] = 1 - 0,8413 = 0,1587 \Rightarrow 15,87\%$$

Ejemplo

Los parámetros de una distribución son $\mu=10$ y desviación típica $\sigma=2$. Se extrae una muestra de 100 individuos. Calcula $P(8 < \bar{X} < 12)$.

Solución:

Por el teorema Central del Límite sabemos que la media muestral de una población normal se distribuye mediante otra distribución normal $N(\mu, \sigma)$ $P(8 < \bar{X} < 12) = \text{tipificando} = P(\frac{8-10}{\frac{2}{10}} < \frac{\bar{X}-10}{\frac{2}{10}} < \frac{12-10}{\frac{2}{10}}) = P(-1 < Z < 1) = 2P(Z < 1) - 1 = 2(0,8416) - 1 = 0,6832$

Ejemplo

Las estaturas, en centímetros, de un grupo de soldados se distribuyen normalmente con media 173 y desviación típica 6.

a) Elegido un soldado al azar, ¿cuál es la probabilidad de que mida menos de 175 cm.?

b) Si se toma una muestra de 12 soldados, ¿cuál es la probabilidad de que su estatura media supere el 176?

Solución:

a) Cuidado, en este apartado no hay muestra, ni por tanto medias muestrales, sino un simple ejercicio de la normal. Si $X = \text{“altura”}$, sabemos que sigue una $N(173,6)$ Luego

$$P(\bar{X} < 175) = \textit{tipificando} = P\left(\frac{\bar{X}-173}{6} < \frac{175-173}{6}\right) = P(Z < 0,33) = 0,6293$$

Aquí si tenemos una muestra con $n=12$

Aunque sea $n < 30$, como la población de partida es normal, podemos también aplicar el Teorema Central del Límite:

$$\bar{X} \rightarrow N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) = N\left(173, \frac{6}{\sqrt{12}}\right) = N(173, 1,73)$$

Luego

$$P(\bar{X} > 176) = \textit{tipificando} = P\left(\frac{\bar{X}-173}{1,73} < \frac{175-173}{1,73}\right) = P(Z > 1,73) = 1 - P(Z < 1,73) = 1 - 0,9582 = 0,0418$$

Ejemplo

Una compañía aérea sabe que el equipaje de sus pasajeros tiene como media 25 kg. con una desviación típica de 6 kg. Si uno de sus aviones transporta a 50 pasajeros, el peso medio de los equipajes de dicho grupo estará en la distribución muestral de medias:

$$N\left(25, \frac{6}{\sqrt{50}}\right) = N(25; 0,84)$$

La probabilidad de que el peso medio para estos pasajeros sea superior a 26 kg sería:

$$P(\bar{x} > 26) = \textit{tipificando} = P\left(\frac{\bar{X}-25}{0,84} > \frac{26-25}{0,84}\right) = P(Z > 1,18) = 1 - P(Z < 1,18) = 0,1190 = 11,90\%$$

Ejemplo

En el control de calidad de una fábrica de latas de atún, se embasan latas de 100 gramos con una desviación típica de 2 gramos. Se empaquetan en cajas de 50 latas. Calcula la probabilidad de que la media de las latas de una caja sea menor que 99 gramos.

Los datos que nos dan son la media poblacional, $\mu=100$, la desviación típica poblacional, $\sigma=2$, y el tamaño de la muestra, $n = 50$.

Sabemos que la media muestral se distribuye según una $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) = N\left(100, \frac{2}{\sqrt{50}}\right) = N(100, 0,28)$

Vamos a recordar como calculábamos esas probabilidades.

Queremos calcular $P(\bar{x} < 99)$.

Lo primero tipificamos para pasar a una distribución $N(0, 1)$.

$$P(\bar{x} < 99) = \textit{tipificando} = P\left(\frac{\bar{X}-100}{0,28} < \frac{99-100}{0,28}\right) = P(Z < -3,54) = 1 - P(Z < 3,54)$$

Recuerda:

La distribución normal es simétrica, por eso en la tabla no aparecen valores negativos, pues los calculamos usando los positivos. Buscamos en la tabla 3'54 y obtenemos que $P(z < 3'54) = 0'9998$.

$P(\bar{x} < 99) = 1 - P(z < 3'54) = 1 - 0'9998 = 0'0002$, una probabilidad muy pequeña.

4.1. Control de la suma de todos los individuos de la muestra

Si tomamos una muestra aleatoria simple de tamaño n de una variable aleatoria $X : X_1, X_2, \dots, X_n$ con media μ y desviación típica σ , entonces la variable aleatoria $X = X_1 + X_2 + \dots + X_n$ para n suficientemente grande ($n \geq 30$) la distribución muestral de medias \bar{X} se aproxima a una distribución normal:

Media: $n\mu$

Desviación típica: $\sigma\sqrt{n}$

$X = N(n\mu, \sigma\sqrt{n})$

Ejemplo

Control de la suma: En el mismo ejemplo anterior determina la probabilidad de que un lote de 400 latas pese más de 40100 gramos.

Como la media muestral es igual a $\bar{x} = \frac{\sum x_i}{n}$ entonces $n\bar{x} = \frac{n \sum x_i}{n} = \sum x_i$ por lo que su distribución es una normal de media $n\mu$ y desviación típica $\sigma = \frac{n\sigma}{\sqrt{n}}$ por lo que será $N(n\mu, \sigma\sqrt{n})$

En nuestro caso $N(n\mu, \sigma\sqrt{n}) = N(400 \times 100, 2\sqrt{400}) = N(40000, 40)$

Queremos calcular

$$P(\sum x_i > 40100) = P(Z > \frac{40100 - 40000}{40}) = P(Z < 2,5) - (1 - P(Z < 2,5)) = 1 - 0,9938 = 0,0062$$

Unas 6 cajas de cada mil pesarán más de 40'1 kg.

Ejemplo

El voltaje de las baterías de un fabricante es de 12 voltios de media y con una desviación típica de 0,3 voltios. Elegimos al azar 5 de esas baterías y se conectan en serie, lo que supone que se suman los voltajes. Calcula la probabilidad de que tengan un voltaje conjunto de más de 61 voltios.

Solución :

El voltaje del conjunto de las baterías sigue una distribución con la siguiente expresión:

$X = \text{sumatorio de las medias de la muestra}$

$$X = N(n\mu, \sigma\sqrt{n}) = N(5 \times 12, 0,3\sqrt{5}) = N(60, 0,67)$$

$$P(\sum x_i > 61) = P(X > 61) = P(Z > \frac{61 - 60}{0,67}) = P(Z > 1,49) = 1 - P(Z < 1,49) = 1 - 0,9319 = 0,0681$$

Ejercicios:

1. Se supone que la distribución de la temperatura del cuerpo humano en la población tiene de media 37° y de desviación típica $0,85^\circ$. Se elige una muestra de 105 personas.
 - a) Hallar las probabilidades de que la media sea menor que $36,9^\circ$
 - b) La media esté comprendida entre $36,5^\circ$ y $37,2^\circ$
2. Las notas de cierto examen se distribuyen según una normal de media 5,8 y desviación típica 2,4. Hallar la probabilidad de que la media de una muestra tomada al azar de 16 estudiantes esté comprendida entre 5 y 7

3. Consideremos la población formada por los cuatro elementos 0, 3, 4, 6. Hallar:
 - a) Todas las muestras posibles de tamaño 2 extraídas mediante muestreo aleatorio simple.
 - b) La media y la desviación típica poblacionales
 - c) La media y la desviación típica de las medias muestrales
4. Los pesos de las ovejas de una cierta ganadería tienen una media de 50 kg con una desviación típica de 4. Elegimos al azar una muestra aleatoria simple de 100 ovejas.
 - a) Determina la probabilidad de que su media sea superior a 51 kg. Sea inferior a 56 kg.
 - b) Sea superior a 48 kg.
 - c) Esté entre 48 kg y 52 kg.
5. Una población tiene una media $\mu = 400$ y una desviación típica = 20. Extraemos una muestra de 1000 individuos.
 - a) Halla el intervalo característico, para una probabilidad de 0'95, de la media muestral.
 - b) Lo mismo para una probabilidad del 0'99.
6. El peso de una población se estima que tiene de media $\mu = 70$ kg y una desviación típica = 10. Se elige una muestra aleatoria simple de 100 individuos y se pesan todos juntos. Calcula la probabilidad de que dicho peso sea superior a 7010 kg.

5. Distribución de las proporciones muestrales

Hemos estudiado el curso pasado la distribución binomial. Era una situación en que las únicas posibilidades eran “éxito” y “no éxito”. Queremos saber cómo se distribuye la proporción muestral (número de éxitos entre el número de veces que se repite el experimento).

Cada muestra que obtengamos de tamaño n se distribuye según una distribución binomial $B(1, p)$, por tanto la suma de n variables $B(1, p)$ es una binomial $B(n, p)$ por el principio de reproductividad de la distribución.

En lugar de calcular medias de las muestras, ahora vamos a trabajar con proporciones.

En una población, la proporción de individuos que poseen una determinada característica es p . (Llamaremos $q = 1 - p$)

Extraemos todas las posibles muestras de tamaño n que podemos extraer de esa población. La proporción de individuos de cada una de esas muestras con esa característica ser \hat{p} .

Llamaremos \hat{P} a la variable aleatoria que toma los distintos valores de esas proporciones muestrales.

Si n es lo suficientemente grande ($n > 30$), se puede demostrar que la variable \hat{P} sigue una distribución normal de parámetros:

$$\hat{P} \Rightarrow N\left(p, \sqrt{\frac{pq}{n}}\right)$$

(Esta fórmula proviene de la aproximación de una binomial por una normal)

Ejemplo 1:

Una nueva droga ha curado al 85% de los enfermos a los que se les ha aplicado. Si se toman muestras de 30 personas, ¿Cuál es la distribución de las proporciones muestrales? ¿Y si las muestras son de 100 personas? ¿Y si son de 1000?

Solución:

Tenemos $p = 0,85$, por lo que $q = 0,15$

Si $n=30$

$$\hat{P} \rightarrow N\left(p, \sqrt{\frac{pq}{n}}\right) = \hat{P} \rightarrow N\left(0,85, \sqrt{\frac{0,85 \cdot 0,15}{30}}\right) = \hat{P} = N(0,85, 0,065)$$

Si $n=100$

$$P \rightarrow \hat{P} \rightarrow N\left(p, \sqrt{\frac{pq}{n}}\right) = P \rightarrow N\left(0,85, \sqrt{\frac{0,85 \cdot 0,15}{100}}\right) = P = N(0,85; 0,036)$$

Si $n=1000$

$$P \rightarrow N\left(p, \sqrt{\frac{pq}{n}}\right) = P \rightarrow N\left(0,85, \sqrt{\frac{0,85 \cdot 0,15}{1000}}\right) = P = N(0,85; 0,011)$$

Como ya vimos con las medias muestrales, al aumentar el tamaño de las muestras disminuye la varianza de las distribuciones muestrales.

Ejemplo

Se sabe que el 15% de los jóvenes entre 18 y 25 años son miopes.

¿Cómo se distribuye la proporción de jóvenes miopes en muestras de 40 individuos? ¿Cuál es la probabilidad de que en dicha muestra la proporción de miopes esté entre el 8 y el 22%? Solucion

Tenemos $p = 0,15$, por lo que $q = 0,85$. Además $n = 40$. Por tanto:

$$P \rightarrow N\left(p, \sqrt{\frac{pq}{n}}\right) = P \rightarrow N\left(0,15; \sqrt{\frac{0,15 \cdot 0,85}{40}}\right) = \hat{P} = N(0,15, 0,0565)$$

Usando el apartado anterior:

$$P(0,08 < \hat{P} < 0,22) = \text{tipificando} = P\left(\frac{0,08-0,15}{0,0565} < Z < \frac{0,22-0,15}{0,0565}\right) = \\ P(-1,24 < Z < 1,24) = 0,785$$

Esto significaría que en el 78,5% de las muestras que extrajésemos de tamaño 40, la proporción de individuos miopes estaría entre el 8 y el 22%.

Ejemplo

El 42% de los habitantes de un municipio es contrario a la gestión del alcalde y el resto son partidarios de este. Si se toma una muestra de 64 individuos, ¿cuál es la probabilidad de que ganen los que se oponen al alcalde?

Solución:

Hay que hallar la probabilidad de que la proporción de la muestra sea superior al 0.5, es decir $\hat{p} = \frac{x}{64} > 0,5$ o equivalentemente, que el número de los que se oponen en la muestra sea $x > 0,5 * 64 = 32$

Otra manera sería. En una muestra de 64 individuos x son los que se oponen.

Para que ganen los que se oponen, x tiene que ser al menos la mitad de los votos mas uno, es decir $x \geq 33$.

$$\text{Siendo } n = 64 \quad \mu = np = 26.88 \quad \sigma = \sqrt{\frac{p \cdot q}{n}} = \sqrt{64 \cdot 0,42 \cdot 0,58} = 3,95$$

Calculamos $P(X \geq 33)$ aproximando a la distribución normal, $x \sim N(26,88; 3,95)$;
 $P(X \geq 33) = 1 - P(X \leq 32) = 1 - P(z < 32,5 - 26,88 / 3,95) = 1 - \Phi(1,42) \approx 0,0778$

EJERCICIOS

- Una máquina fabrica piezas de precisión y en su producción habitual tiene un 3% de piezas defectuosas. Se empaquetan en cajas de 200, ¿cuál es la probabilidad de encontrar entre 5 y 7 piezas defectuosas en una caja?
- Si tiramos una moneda no trucada 100 veces, ¿cuál es la probabilidad de que obtengamos más de 55 caras?
- Se ha realizado una encuesta a 20.000 universitarios sobre su actitud ante el botellón. De ellos, 13.200 son partidarios y el resto no. Se toman 100 muestras de 30 universitarios cada una. Hallar:
 - La distribución de la proporción muestral de partidarios del botellón
 - La probabilidad de que en una de estas muestras se manifiesten favorables al botellón más de 21 alumnos
 - ¿En cuántas muestras cabe esperar que más de 15 y menos de 19 estudiantes se muestren partidarios del botellón?
- En el instituto de cierta localidad se imparten 4 niveles diferentes: 1º, 2º, 3º y 4º ESO, estando matriculados un total de 800 alumnos. En 1º hay 160 alumnos, en 2º hay 240 y en 3º hay 208. Explica cómo obtener una muestra de 50 alumnos mediante:
 - Muestreo aleatorio sistemático
 - Muestreo aleatorio estratificado con afijación proporcional
- En un barrio hay 4000 habitantes, distribuidos en cuatro urbanizaciones: el 12% viven en A, el 20% en B, el 36% en C y el 32% en D. Además, los porcentajes de mujeres en cada urbanización son: 50, 60, 66 y 75% respectivamente. Se desea obtener una muestra de 50 habitantes que sea representativa en cuanto al sexo y a las urbanizaciones mediante muestreo aleatorio estratificado proporcional. ¿Cuántas personas (hombres y mujeres) de cada urbanización habrá que seleccionar?
- Dada la población {2, 4, 6, 8}: Escribir todas las muestras posibles de tamaño 2 mediante muestreo aleatorio simple. Calcular la media y la varianza de las medias muestrales. Hacer lo mismo con todas las muestras posibles de tamaño 3

7. La media de edad de los lectores de una determinada revista es de 17'2 años y la desviación típica de 2'3 años. Si elegimos muestras de 100 lectores: ¿Cuál es la distribución de las medias muestrales? ¿Cuál es la probabilidad de que la media de edad de la muestra esté comprendida entre 16'7 y 17'5 años?
8. En una determinada población, los pesos se distribuyen según una normal de media 65 kg. y varianza 49. Si extraemos muestras de tamaño 64, ¿cuál es la probabilidad de que la media de los pesos de una de esas muestras sea mayor que 66'5 kg?
9. La edad de los miembros de una determinada asociación se distribuye normalmente con media 52 años y desviación típica 3 años. ¿Cuál es la probabilidad de que un miembro elegido al azar sea mayor de 60 años? Si se toman muestras de tamaño 36, ¿Cuál es la probabilidad de que la edad media de dichas muestras sea superior a los 60 años?
10. La edad de los alumnos de 2º Bachillerato de cierto instituto sigue una distribución $N(17'6, 0'5)$. Los agrupamos al azar de 10 en 10 para una competición. ¿Cuál es la probabilidad de que en uno de esos grupos la edad media esté comprendida en el intervalo (17,18)?