Unidad 5.- TEORÍA DE MUESTRAS

CONCEPTOS GENERALES

Se llama ***POBLACIÓN*** al conjunto de todos los elementos que poseen una determinada característica. Los distintos elementos de la población se llaman ***INDIVIDUOS***.

Cada una de las características poblacionales se llama variable.

Las variables pueden ser de 2 tipos:

- ***CUALITATIVA***, cuando describen una **característica no numérica** de la población.
- ***CUANTITATIVA***, si dicha **característica es numérica** y pueden ser discretas y continuas

Se denomina ***MUESTRA*** a la parte de la población que se observa para extraer información sobre la población.

Se llama ***MUESTREO*** al proceso mediante el cual se selecciona la muestra de la población.

Representatividad de la muestra

A la hora de elegir la muestra hay dos aspectos que son los que van a determinar si los resultados del estudio estadísticos son válidos o no: el tamaño de la muestra y la forma de seleccionarla.

Tamaño de la muestra

Si el tamaño de la muestra es demasiado pequeño, los resultados son poco fiables.

Si el tamaño de la muestra es elevado, los resultados son más fiables, pero un tamaño demasiado elevado conlleva un mayor gasto económico y laboral.

El objetivo es **mantener un equilibrio entre los costes y la validez de los resultados obtenidos.**

Forma de seleccionar la muestra

La muestra seleccionada debe ser ***representativa*** de la población estudiada, ya que si no, las conclusiones obtenidas en el estudio no serán fiables. Es necesario que sus elementos hayan sido ***seleccionados de manera aleatoria***.

TIPOS DE MUESTREO

Muestreo aleatorio simple

Se dice que una muestra fue obtenida por muestreo aleatorio simple (o simplemente que es una muestra aleatoria simple) si todos los individuos de la población tienen la misma probabilidad de salir elegidos en la muestra.

Ejemplo: Se quiere averiguar el gasto medio de las personas que acuden a un centro comercial

Se elige al azar a 100 personas que hayan acudido a ese centro.

Muestreo aleatorio sistemático

Es el que consiste en obtener el 1er individuo al azar y después tomar los siguientes "a saltos" de magnitud igual dentro de la lista.

Ejemplo: En una etapa de una vuelta ciclista se quiere realizar un control antidopaje.

Elegimos al azar un ciclista al llegar a la meta y se le hace pasar el control. A continuación se selecciona *de k en k* a los siguientes corredores a medida que pasan por la llegada.

Muestreo aleatorio estratificado

Es una técnica muy usada cuando la población puede no ser tan homogénea como se desearía y existen distintos grupos (llamados estratos) en los que la variable en estudio puede diferir bastante de unos a otros. Por ejemplo pueden estar delimitados por la edad o el sexo.

En este caso, la población se divide en grupos homogéneos (estratos) y posteriormente se extrae una muestra aleatoria de cada estrato de forma que en la muestra, cada estrato mantenga la misma proporción que en la población

Ejemplo: El 60% de los estudiantes de un colegio son chicos y el 40%,chicas. ¿Cómo habría que seleccionar una muestra de 80 alumnos?

Si se quiere elegir una muestra de 80 alumnos, el nº de alumnos y alumnas debe elegirse de manera proporcional, es decir, debe haber 60% de chicos y 40% de chicas, es decir se elegirán aleatoriamente 48 alumnos y 32 alumnas.

Muestreo aleatorio por conglomerados

Supone que la población se compone de grupos, pero que ahora son homogéneos entre sí (a diferencia de los estratos). En este caso se eligen (por motivos económicos) algunos de los conglomerados de los que se obtiene la muestra

Ejemplo: Si se quiere obtener una muestra de la población gallega de más de 18 años con el objeto de estudiar el color de los ojos

Se pueden tomar como conglomerados los distintos ayuntamientos de Galicia y así poder tomar como muestra elegidos aleatoriamente unos cuantos ayuntamientos.

ESTIMACIÓN PUNTUAL

El objetivo de la ***estimación puntual*** es *obtener un valor numérico* para cierto parámetro de la población a estudio.

- Así, si hacemos una encuesta y estamos interesados en conocer el apoyo a un candidato a la alcaldía, debemos estimar la **PROPORCIÓN** de votos que obtendrá dicho candidato.
- Si lo que pretendemos es averiguar cuál es el número medio diario de usuarios de un determinado servicio, el parámetro que debemos estimar será la **MEDIA** poblacional.

Dado un parámetro, θ, de una población y una muestra obtenida de la misma, un estimador no es más que una expresión, $\hat{\theta}$, que se calcula con dicha muestra y que está destinada a obtener un valor próximo al del parámetro desconocido.

Ejemplo: Tomando como parámetro la estatura media, un posible estimador sería la media muestral (media aritmética de los valores de las estaturas de la muestra y divididos entre el tamaño muestral)

DISRIBUCIÓN EN EL MUESTREO DE UNA PROPORCIÓN

Ejemplo. Una empresa que fabrica CD tiene la sospecha de que están produciendo CD defectuosos. La empresa quiere averiguar la proporción, p, de CD que funcionan correctamente.

Sea p la proporción de CD que funcionan correctamente. Se desconoce su valor, pero existen métodos para aproximarlo.
Se toma una muestra aleatoria de 500 CD y se comprueba que 430 de ellos están en buen estado.
El valor 430/500 se denota por $\hat{p}$ e indica la proporción de CD en buen estado que hay en la muestra seleccionada, pero no coincide con p. es obvio que si se hubiese elegido otra muestra de 500 CD, el valor de $\hat{p}$ sería diferente.
Los distintos valores de $\hat{p}$ que dependerán de las muestras elegidas, dan lugar a una variable aleatoria que se representa por $\hat{P}$ y que se llama **estadístico.**
La distribución de los valores de $\hat{P}$ se llama **distribución en el muestreo de una proporción.**

La variable aleatoria $\hat{p}$ tiene las siguientes características:
1° Media: $\mu = p$
2° Desviación típica $\sigma = \sqrt{\frac{p(1-p)}{n}}$
3° A medida que n crece, la distribución de $\hat{p}$ se aproxima cada vez más a una normal $N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$, siempre que p no se acerque ni a 0 ni a 1.

Resumen:

$$\hat{p} : N\left(p\,, \sqrt{\frac{p(1-p)}{n}}\right)$$

DISTRIBUCIÓN EN EL MUESTREO DE LA MEDIA

Ejemplo. La dirección General de Tráfico desea conocer el peso medio de la carga que transportan los turismos que circulan por las carreteras en una operación salida.

La carga media de los turismos se denota por μ y la desviación típica por σ.

Unidad 5.- TEORÍA DE MUESTRAS

Con el fin de tener una idea aproximada de cuánto puede valer μ se elige una muestra aleatoria de 100 turismos y se obtiene que:

1º el peso medio muestral es $\overline{x}_1 = 68{,}5\ kg$

2º La desviación típica de la muestra es s_1 = 12,6 kg

Si se eligen otras muestras de tamaño 100 y se calculan sus medias y desviaciones típicas, se obtendrán otras medias, $\overline{x}_2, \overline{x}_3, \ldots$ y otras desviaciones típicas, s_2, s_3,....

Los diferentes valores de $\overline{x}_i$ dan lugar a una variable aleatoria que se representará por $\overline{X}$ y que se denomina **estadístico.**

La distribución de los valores de $\overline{X}$ se llama **distribución en el muestreo de la media**

La variable aleatoria $\overline{X}$ tiene las siguientes características:

1º Media μ (la misma que la de la población)

2º Desviación típica $\frac{\sigma}{\sqrt{n}}$

3º A medida que n crece, la distribución de se aproxima cada vez más a una normal $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

Resumen:

$$\overline{X} : N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

TEOREMA CENTRAL DEL LÍMITE

Sea X una variable aleatoria de una población de media μ y desviación típica σ. entonces se verifica que.

1º La distribución de las medias muestrales de tamaño n tiene media μ y desviación típica $\frac{\sigma}{\sqrt{n}}$

2º La distribución de las medias muestrales se aproxima a una normal a medida que crece el tamaño de la muestra

Este resultado se conoce como teorema central del límite y fue enunciado por Laplace.

* Si la **población** de partida **es normal** la **distribución** de las medias **será normal** independientemente del tamaño de la muestra.

* Si la **población** de partida **no es normal** la distribución de las medias se podrá aproximar por una normal cuando el **TAMAÑO** de la muestra sea **mayor o igual a 30** $(n \geq 30)$

PROBLEMAS RESUELTOS

1.- Ante una medida adoptada por el Gobierno, se sabe que el 35% de la población está a favor de dicha medida. Se toma una muestra al azar de 200 personas.

a) Qué número de personas estén a favor?

b) Halla la distribución de muestreo de la proporción de personas que están a favor de la medida.

c) Halla la probabilidad de que en la muestra elegida, más de la mitad de los integrantes que la forman estén a favor.

Solución:

a) El número esperado será $\mu = n \cdot p$; $\mu = 200 \cdot 0{,}35 = 70$ personas

b) La variable aleatoria es la proporción $\hat{p}$ que sigue una distribución normal

$$N\left(p\,,\sqrt{\frac{p(1-p)}{n}}\right) = N\left(0{,}35\,,\sqrt{\frac{0{,}35\cdot 0{,}65}{200}}\right) = N(0{,}35\,;0{,}034)$$

c) $P(\hat{P} > 0{,}5) = P\left(Z > \frac{0{,}5-0{,}35}{0{,}034}\right) = P(Z > 4{,}41) \approx 0$

Es decir, es prácticamente imposible que más de la mitad de los integrantes de la muestra estén a favor de la media.

2.- La emisión de óxido de nitrógeno de los vehículos de cierta marca sigue una distribución normal con media 1,2 y desviación típica 0,4. Se escoge al azar una muestra de 25 vehículos.

a) ¿Cuál es la distribución en el muestreo de la media?

b) Halla la probabilidad de que la media de la muestra sea mayor de 1,2.

Solución:

a) $\overline{X}$ sigue una normal

$$\overline{X} : N\left(\mu\,,\frac{\sigma}{\sqrt{n}}\right) = N\left(1{,}2\,;\frac{0{,}4}{\sqrt{25}}\right) = N\,(1{,}2\,;\ 0{,}08)$$

b) $P(\overline{X} > 1{,}2) = P\left(Z > \frac{1{,}2-1{,}2}{0{,}08}\right) = P(Z > 0) = 1 - P(Z \leq 0) = 1 - 0{,}5 = 0{,}5$