TEORÍA DE MUESTRAS

Índice:

1. Introducción

En el estudio de los fenómenos aleatorios, lo que interesa, es tener el mayor conocimiento posible, utilizando los mínimos recursos posibles. Por ejemplo, si un determinado país, quiere tener una cantidad aconsejable de vacunas (*que sobren o falten las menos posibles*) para prevenir de algún tipo de enfermedad, tendrá que prever mediante alguna técnica estadística cuantas vacunas necesitará.

La **teoría de muestras** estudia las técnicas y procedimientos que debemos emplear para que las muestras sean representativas de la población que pretendemos estudiar, de forma que los errores en la determinación de los parámetros de la población objeto de estudio sean mínimos.

2. Muestra y Población.

Población: Es el conjunto de todos los elementos que poseen una determinada característica. En general, supondremos que la población es muy grande.

Muestra: Es un subconjunto de la población.

Muestreo: Es el proceso mediante el cual se escoge una muestra de la población

 # Ejemplo.- si queremos conocer la cantidad de tornillos que se fabrican diariamente en una fabrica de utillaje, lo conveniente sería coger una muestra de tornillos, para deducir la cantidad de tornillos defectuosos. Sin embargo, para que el estudio sea efectivo deberá ser elegido de forma aleatoria, pues en otro caso las conclusiones no serían fiables.

3. Tipo de muestreos

Los distintos tipos de muestreo que podemos utilizar en un estudio estadístico son muchos, además, de los que podamos inventar. Algunos tipos de muestreos más utilizado son

Muestreo aleatorio simple.- Es el muestreo en cual todos los elementos de la población tienen la misma probabilidad de ser elegidos para formar parte de de la muestra.

 # Ejemplo.- Si queremos conocer los preferencias deportivas de una determinada ciudad, podemos efectuar una muestra aleatoria simple.

Muestreo aleatorio estratificado.- Es el muestreo que previamente divide la población en grupos homogéneos, denominados estratos, y posteriormente se utiliza muestras aleatorias de dichos grupos o estratos.

 # Ejemplo.- Si queremos conocer la profesión de una determinada población según el sexo, o el estado civil, es conveniente utilizar un muestreo estratificado.

Muestreo aleatorio sistemático.- Es el muestreo que selecciona al azar un elemento de la población y a partir de él se seleccionan de k en k los siguientes.

 # Ejemplo.- Si en un determinado municipio, se quiere estimar la cantidad de gente que fuma, y se puede tomar una muestra aleatoria sistemática a partir del censo de la población.

Muestreo por conglomerados.- Es el muestreo que previamente divide la población en secciones o conglomerados, se eligen al azar una muestra de estas secciones, y se toman todos los elementos de la secciones elegidas para formar la muestra.

 # Ejemplo.- Si queremos conocer las preferencias de los alumnos y alumnos de los Institutos de Castilla la Mancha, podemos tomar un muestreo aleatorio por conglomerados, tomando una muestra aleatoria de todos los Institutos de Castilla la Mancha, y luego preguntar a todos los alumnos de dichos centros elegidos.

4. Distribución en el muestreo de una proporción

Si $\hat{p}$ es el **estadístico** (*proporción de una muestra aleatoria*) de la proporción p de una población. A la distribución de probabilidad de $\hat{p}$ se le denomina **distribución muestral o distribución en el muestreo de la proporción** $\hat{p}$, y cumple:

1.- La media es $\mu = p$

2.- La desviación típica es $\sigma = \sqrt{\frac{p.(1-p)}{n}}$

3.- Cuando n crece, la distribución de probabilidad de $\hat{p}$ se acerca a al distribución de probabilidad $N(\mu, \sigma)$ (*siempre que p no se aproxime ni a 0 , ni a 1*).

 # Ejemplo.- Un nuevo medicamento ha curado al 85% de los enfermos a los que se les ha suministrado. Podemos determinara la proporción de enfermos curados para muestras de tamaño 30, 100 y 1000 personas.

Teniendo en cuenta que la proporción de enfermos es $p = 0{,}85$ *, podemos construir la tabla, que determina las respectivas distribuciones muestrales según el tamaño de la muestra*

Tamaño de la muestra n	Media $\mu = p$	Desviación típica $\sigma = \sqrt{\frac{p.(1-p)}{n}}$	Distribución normal $N(\mu, \sigma)$
30	0.85	$\sqrt{\frac{0{,}85 . 0{,}15}{30}} = 0{,}065$	$N(0{,}85 ; 0{,}065)$
100	0.85	$\sqrt{\frac{0{,}85 . 0{,}15}{100}} = 0{,}036$	$N(0{,}85 ; 0{,}036)$
1000	0.85	$\sqrt{\frac{0{,}85 . 0{,}15}{1000}} = 0{,}011$	$N(0{,}85 ; 0{,}011)$

5. Distribución en el muestreo de la media

Si $\bar{X}$ es el **estadístico** (*promedio de una muestra aleatoria de tamaño n*) de la media μ, de una población de la que conocemos la desviación típica σ. A la distribución de probabilidad de $\bar{X}$ se le denomina **distribución de las medias muestrales o distribución en el muestreo de la media**, $\bar{X}$ y cumple:

1.- La media es μ

2.- La desviación típica es $\frac{\sigma}{\sqrt{n}}$

3.- Cuando n crece, la distribución de probabilidad de $\bar{X}$ se acerca a al distribución de probabilidad $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$.

Si no conocemos σ y $n \geq 30$, podemos sustituir σ por $\hat{s} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})$

 # Ejemplo.- Se supone que la distribución de la temperatura del cuerpo humano en la población tiene de media 37º y de desviación típica 0,85º. Si se elige una muestra de 105 personas.

Como la distribución de la media muestral $\bar{X}$ *sigue una distribución*

$$N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) = N\left(37, \frac{0{,}85}{\sqrt{105}}\right) = N(37; 0{,}829) \quad .$$

Si p es la función de probabilidad de dicha distribución, teniendo en cuenta que la variable

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{X} - 37}{0{,}0829}$$

sigue una distribución $N(0,1)$ *, se cumplirá*

$$p(\bar{X} \leq 36{,}9) = p\left(Z \leq \frac{36{,}9 - 37}{0{,}0829}\right) = p\left(Z \leq \frac{36{,}9 - 37}{0{,}0829}\right) = p(Z \leq -1{,}205) = p(Z \geq 1{,}205) =$$
$$= 1 - p(Z \leq 1{,}205) = 1 - 0{,}8849 = 0{,}115$$

$$p(36{,}5 < \bar{X} \leq 37{,}5) = p([36{,}5 < \bar{X} \leq 37] \cup [37 < \bar{X} \leq 37{,}5]) = 2.\, p(37 < \bar{X} \leq 37{,}5) =$$
$$= 2.\, p\left(0 < Z \leq \frac{37{,}5 - 37}{0{,}0829}\right) = 2.\, p(0 < Z \leq 6{,}031) = 2.[\, p(Z \leq 6{,}031) - p(Z > 0)] =$$
$$= 2.[\, p(Z \leq 6{,}031) - (1 - p(Z \leq 0))] = 2.[0{,}9999\ldots - (1 - 0{,}5)] \approx 1$$

6. Distribución de las sumas muestrales

Si elegimos n muestras del mismo tamaño y calculamos sus medias $t_1, t_2, \ldots, t_n$ y sus desviaciones típicas $s_1, s_2, \ldots, s_n$ de una población de media μ y de desviación típica σ

La suma de los distintos valores t_i dan lugar a una variable aleatoria que representamos por T, y cuya distribución se le denomina **distribución de las sumas muestrales o distribución en el muestreo de las sumas** , y cumple:

1.- La media es $n\mu$

2.- La desviación típica es $\sigma.\sqrt{n}$

3.- Cuando n crece, la distribución de probabilidad de $T=t_1+t_2+\ldots+t_n$ se acerca a la distribución de probabilidad $N\left(n\mu, \sigma.\sqrt{n}\right)$.

 # Ejemplo.- Consideremos la población formada por 3 bolas contenidas en una urna numeradas del 1 al 3.

La distribución de las sumas muestrales de 2 muestras extraídas con devolución, como la media y la desviación típica de la población son respectivamente

$$\mu=\frac{1+2+3}{3}=2 \qquad \sigma=\sqrt{\frac{(1-2)^2+(2-2)^2+(3-2)^2}{3}}=\sqrt{\frac{2}{3}}=\frac{\sqrt{6}}{3}$$

Entonces, la distribución de las sumas de las medias muestrales de T es

$$N(n\mu, \sigma.\sqrt{n})=N\left(2.2; \frac{\sqrt{6}}{3}.\sqrt{2}\right)=N\left(4; \frac{2.\sqrt{3}}{3}\right)$$

7. Distribución de las diferencia de medias muestrales

Si elegimos dos muestras de $\bar{X}_1$ y $\bar{X}_2$ de tamaño n de dos poblaciones de medias μ_1 y μ_2 , y desviaciones típicas σ_1 y σ_2 respectivamente.

Los distintos valores $\bar{x}_i-\bar{x}_2$ dan lugar a una variable aleatoria que representamos por $\bar{X}_i-\bar{X}_2$, y cuya distribución se le denomina **distribución de las diferencias de las medias muestrales**, y cumple:

1.- La media es $\mu_1-\mu_2$

2.- La desviación típica es $\sqrt{\left(\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)}$

3.- Cuando n_1 y n_2 crece, la distribución de probabilidad de $\bar{X}_i-\bar{X}_2$ se acerca a la distribución de probabilidad $N\left(\mu_1-\mu_2, \sqrt{\left(\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)}\right)$.

Si las desviaciones típicas son desconocidas y las muestras son grandes, sustituimos las desviaciones típicas σ_1 y σ_2 por las desviaciones muestrales $\hat{s}_1$ y $\hat{s}_2$.

 # Ejemplo.- Se recoge al azar una muestra de 40 hombres, los cuales tiene un salario medio de 1520 € con una desviación típica muestral de 70 €. También se escoge al azar una muestra de 30 mujeres que tiene un salario medio de 1470 € con desviación típica muestral de 50 €. ¿Cuál es la probabilidad de que la diferencia de los sueldos medios sea mayor que 60 €?

Teniendo en cuenta que las respectivas medias y desviaciones típicas de las dos muestras de salarios es

$\bar{x}_1=1520€$; $\hat{s}_1=70€$; $n_1=40$ *(hombres)*

$\bar{x}_2=1470€$; $\hat{s}_2=50€$; $n_2=30$ *(mujeres)*

Como la variable $\bar{X}_1-\bar{X}_2$ *, sigue una distribución*

$$N\left(1520-1470, \sqrt{\frac{70^2}{40}+\frac{50^2}{30}}\right)=N\left(50, \sqrt{\frac{1235}{6}}\right)\approx N(50, 14'35)$$

Luego, si p es la función de probabilidad se cumplirá

$$p(\bar{X}_1-\bar{X}_2>60)=p\left(\bar{Z}_1-\bar{Z}_2>\frac{60-50}{14'35}\right)=p(\bar{Z}_1-\bar{Z}_2>0{,}70)=$$
$$=1-p(\bar{Z}_1-\bar{Z}_2\leq 0{,}70)=1-0{,}7580=0{,}242$$

8. Teorema central del límite

sea X una variable aleatoria de una población de media μ y desviación típica σ , entonces se verifica que:

1. La distribución de las medias muestrales de tamaño n, tiene media μ y desviación típica $\frac{\sigma}{\sqrt{n}}$
2. La distribución de las medias muestrales se aproxima a una normal a medida que crece el tamaño de la muestra n.

 # Ejemplo.- El cociente intelectual de unos universitarios se distribuye normalmente con media 100 y desviación típica 11.

a) Se elige una persona al azar. Hallar la probabilidad de que su CI esté entre 100 y 103.

b) Se elige al azar una muestra de 25 personas. Encontrar la probabilidad de que la media de los cocientes intelectuales esté entre 100 y 103.

a) Si X es la variable aleatoria que define el coeficiente intelectual, como X sigue la variable aleatoria N(100,11). Si p es la probabilidad, será

$$p(100<X\leq 103)=p\left(\frac{100-100}{11}<Z\leq\frac{103-100}{11}\right)=p(0<Z\leq 0{,}27)=$$
$$=p(Z\leq 0{,}27)-p(Z\leq 0)=0{,}6064-0{,}5=0{,}1064$$

b) Como la distribución de la media muestral que procede de una distribución normal, es normal cualquiera que sea el valor n, y se distribuirá según una distribución

$$N\left(\mu,\frac{\sigma}{\sqrt{n}}\right)=N\left(100,\frac{11}{\sqrt{25}}\right)=N(100{,}2{,}2)$$

Luego, se cumplirá

$$p(100<\bar{X}\leq 103)=p\left(\frac{100-100}{2{,}2}<Z\leq\frac{103-100}{2{,}2}\right)=p(0<Z\leq 1{,}36)=$$
$$=p(Z\leq 1{,}36)-p(Z\leq 0)=0{,}9131-0{,}5=0{,}4131$$