INFO@ADEFACIL.COM
ELENA GONZALO
671 227 871

ÍNDICE

INFO@ADEFACIL.COM
ELENA GONZALO
671 227 871

T5. ESTADÍSTICOS MUESTRALES Y SUS DISTRIBUCIONES

5.1 INTRODUCCIÓN. MUESTRA ALEATORIA

A veces se recurre a la selección de una muestra, una parte de la población, con el fin de extrapolar la información que nos proporcionen los elementos de la muestra a toda la población mediante diferentes técnicas estadísticas. Este proceso recibe el nombre de ***inferencia estadística***.

Para poder asegurar la *calidad* de la inferencia, la muestra debe ser representativa de la población y sus elementos no deben reunir ninguna característica esencial que los diferencie del resto. La *representatividad* de la muestra se trata de conseguir seleccionando las unidades de manera aleatoria.

Existen dos tipos de muestreo aleatorio:

- ***Muestreo con reemplazamiento***. Cada unidad seleccionada se devuelve a la población pudiendo resultar elegida nuevamente. Las probabilidades sucesivas de selección de un elemento en cada una de las n extracciones muestrales serán:

$$\frac{1}{N}, \frac{1}{N}, \dots, \frac{1}{N}$$

- ***Muestreo sin reemplazamiento.*** La unidad seleccionada se retira de la población no pudiendo ser elegida otra vez. Las probabilidades sucesivas de selección dependerán en cada extracción de la anterior:

$$\frac{1}{N}, \frac{1}{N-1}, \dots, \frac{1}{N-n+1}$$

Si la población es muy grande o infinita, entonces no existe una diferencia significativa.

A partir de ahora nos referiremos a poblaciones de tamaño infinito en las que no haremos referencia a que el muestreo sea con o sin reposición.

Sea X la variable aleatoria correspondiente a una población con función de distribución $F(x)$. Silas variables aleatorias X_1, X_2, …, X_n son independientes y tienen la misma función de distribución, $F(x)$, que la de la distribución de la población, entonces las variables aleatorias X_1, X_2, …, X_n forman un conjunto de variables aleatorias independientes e idénticamente distribuidas que constituyen una muestra aleatoria simple de tamaño n de la población $F(x)^2$.
Si la población de partida es de tipo discreto, entonces la función de probabilidad conjunta de la muestra será:

$$P(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^{n} P(X = x_i) = \prod_{i=1}^{n} p_i = p_1 \cdot \dots \cdot p_n$$

Si la población es de tipo continuo con función de densidad $f(x)$, entonces la función de densidad de la muestra será:

$$f(x_1, \dots, x_n) = \prod_{i=1}^{n} f(x_i) = f(x_1) \cdot \dots \cdot f(x_n)$$

5.2 PARÁMETROS POBLACIONALES Y ESTADÍSTICOS MUESTRALES

Los parámetros poblacionales son las características numéricas de la distribución poblacional. Si la característica sigue una distribución normal, $N(\mu, \sigma)$, observamos que para describir totalmente la función de densidad tendremos que dar valores a los dos.

Se presenta un problema cuando deseamos estudiar una población con función de distribución $F(x,\theta)$, donde la forma de la distribución es conocida pero depende de un parámetro θ desconocido.

Un ***estadístico***, T, es cualquier función real de las variables aleatorias que integran la muestra, la cual no contiene ningún valor o parámetro desconocido:

$T = (X_1, \dots , X_n)$

Para una muestra concreta $(x_1, \dots , x_n)$ el estadístico (usaremos letras mayúsculas para indicar las variables aleatorias y letras minúsculas para los valores concretos) tomará el valor $T = g(x_1, \dots ,x_n)$ y a medida que vamos tomando muestras diferentes se obtienen distintos valores del estadístico, por lo tanto, el estadístico T es también una variable aleatoria que tendrá su distribución probabilística, a la que llamaremos ***distribución muestral del estadístico***.

El parámetro es una constante y cuando se conoce determina completamente el modelo probabilístico, sin embargo el estadístico es una variable aleatoria cuyo valor dependerá de las observaciones muestrales.

Si la medida numérica (media, desviación típica, etc.) se calcula para el conjunto de datos poblacionales le llamaremos valor del parámetro poblacional y si se calcula para el conjunto de datos muestrales, le llamaremos valor del estadístico muestral.

En una población finita de tamaño N los parámetros poblacionales media, varianza y proporción poblacional son:

$$\boldsymbol{\mu} = \frac{1}{N} \cdot \sum_{i=1}^{N} X_i \qquad \boldsymbol{\sigma^2} = \frac{1}{N} \cdot \sum_{i=1}^{N} (X_i - \mu)^2 \qquad \boldsymbol{p} = \frac{X}{N} = \frac{n^{\underline{o}}\ de\ éxitos\ en\ N\ pruebas}{n^{\underline{o}}\ de\ pruebas}$$

INFO@ADEFACIL.COM
ELENA GONZALO
671 227 871

Para una muestra aleatoria simple de tamaño n, $(X_1, \ldots, X_n)$ los estadísticos media, varianza y proporción muestral son:

$$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i \qquad S^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (X_i - \bar{X})^2 \qquad P_X = \frac{X}{N} = \frac{n^{\underline{o}}\ de\ éxitos\ en\ n\ pruebas}{n^{\underline{o}}\ de\ pruebas}$$

El estadístico varianza muestral, s^2, se puede formular también como:

$$S^2 = \frac{1}{n-1} \cdot \left(\sum_{i=1}^{n} X_i^2 - n \cdot \bar{X}^2 \right) = \frac{1}{n-1} \cdot \left(\sum_{i=1}^{n} X_i^2 - \frac{(\sum_{i=1}^{n} X_i)^2}{n} \right)$$

5.3 FUNCIÓN DE DISTRIBUCIÓN EMPÍRICA

La función de distribución de una variable aleatoria X estaba definida como:

$$F(x) = P(X \leq x)$$

Que representa la proporción de valores que son menores o iguales que x.

Consideremos una población con función de distribución $F(x)$ y sean $(x_1, \ldots, x_n)$ los valores observados en la muestra correspondientes a una muestra aleatoria simple procedente de esa población, y designamos por $N(x)$ el número de valores observados que son menores o iguales que x. Entonces definimos la ***función de distribución empírica de la muestra*** como:

$$F_n(x) = \frac{N(x)}{n}$$

Un *estadístico* es una variable aleatoria que tendrá su distribución de probabilidad denominada ***distribución muestral*** (hay diferencia entre la distribución de la población de la cual se ha extraído la muestra y la distribución de alguna función de esa muestra, el estadístico).

Una vez obtenida la distribución muestral de un estadístico se podrán hacer afirmaciones probabilísticas sobre él y estudiar sus propiedades como estimador.

INFO@ADEFACIL.COM
ELENA GONZALO
671 227 871

5.4 DISTRIBUCIÓN MUESTRAL DE ESTADÍSTICOS

Un *estadístico* es una variable aleatoria que tendrá su distribución de probabilidad denominada ***distribución muestral*** (hay diferencia entre la distribución de la población de la cual se ha extraído la muestra y la distribución de alguna función de esa muestra, el estadístico).

Una vez obtenida la distribución muestral de un estadístico se podrán hacer afirmaciones probabilísticas sobre él y estudiar sus propiedades como estimador.

Teorema

Si $(X_1,\dots, X_n)$ es una muestra aleatoria simple de tamaño n procedente de una población, descrita por la variable aleatoria X, con media $E[X] = \mu$ y varianza $Var(X) = \sigma^2$, entonces la *esperanza de la media muestral* es igual a la media de la población, μ, y la *varianza de la media muestral* es igual a la varianza poblacional, σ^2, dividida por n.

$$E[\bar{X}] = \mu \qquad\qquad Var(\bar{X}) = \frac{\sigma^2}{n}$$

A la correspondiente *desviación típica* del estadístico $\bar{X}$ se le llama ***error estándar de la media*** y viene dado por:

$$\boldsymbol{\sigma(\bar{X})} = error\ estándar\ de\ la\ media\ muestral\ \bar{X} = \frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{n}}}$$

El valor central del estadístico media muestral es la media poblacional μ, resulta que cuanto mayor sea el tamaño muestral n menor será la dispersión de $\bar{X}$ en torno a la media poblacional μ, y el valor observado del estadístico estará más próximo a μ. El estadístico media muestral $\bar{X}$ puede ser considerado como un buen estimador de la media poblacional μ.

Para hacer inferencias sobre el parámetro media poblacional μ utilizando $\bar{X}$ no es necesario tomar muestras de tamaño demasiado grande.

El Teorema anterior es válido cuando el muestreo se hace de una población infinita, o de una finita, pero con reemplazamiento. Si el muestreo se hace sin reemplazamiento en una población finita de tamaño N, tendríamos que:

$$E[\bar{X}] = \mu \qquad\qquad Var(\bar{X}) = \frac{\sigma^2}{n}\cdot\boldsymbol{\frac{N-n}{N-1}}$$

Al término en negrita se le suele llamar ***factor de corrección de población finita***.

Teorema

Si $(X_1, \ldots, X_n)$ es una muestra aleatoria simple de tamaño n, procedente de una población, descrita por la variable aleatoria X, con varianza $Var(X) = \sigma^2$, entonces la *esperanza de la varianza muestral* S^2 es igual a la varianza poblacional σ^2 y la varianza de la varianza muestral es función del momento central de orden cuatro:

$$E[S^2] = \sigma^2 \qquad Var(S^2) = \frac{\mu_4}{n} + \frac{3-n}{n \cdot (n-1)} \cdot \sigma^4$$

5.5 DISTRIBUCIÓN MUESTRAL DE POBLACIONES NORMALES

5.5.1 Distribución de la media muestral cuando se conoce la varianza poblacional.

Sea $(X_1, \ldots , X_n)$ una muestra aleatoria simple de tamaño n, procedente de una población $N(\mu, \sigma)$. Entonces la distribución del estadístico media muestral tendrá una distribución normal:

$$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i \to N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Y, como consecuencia, el estadístico tipificado seguirá una normal estándar:

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \to N(0,1)$$

En muchas situaciones la población de partida de la que se extrae la muestra no es normal. En tales casos, la distribución muestral asintótica del estadístico media muestral $\bar{X}$, para muestras grandes ($n \geq 30$) seguirá siendo normal.

Teorema

Sea $(X_1,\ldots, X_n)$ una muestra aleatoria simple de tamaño n, procedente de una población $N(\mu,\sigma)$. Entonces las variables aleatorias:

$$Z_i = \frac{X_i - \mu}{\sigma}$$

Son $N(0,1)$ e independientes y tales que

$$\sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2$$

Sigue una distribución χ^2 con n grados de libertad.

INFO@ADEFACIL.COM
ELENA GONZALO
671 227 871

5.5.2 Distribución de la media muestral cuando no se conoce la varianza poblacional.

Cuando la varianza poblacional es desconocida se procede como indica el siguiente teorema y consiste en estimar la varianza poblacional σ^2 utilizando la varianza muestral S^2.

Teorema

Si $(X_1, \dots, X_n)$ es una muestra aleatoria simple de tamaño n, procedente de una población $N(\mu, \sigma)$ con σ desconocida, entonces el estadístico

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \rightarrow t - Student\ con\ (n-1)\ grados\ de\ libertad.$$

Si la muestra es grande, la distribución del estadístico anterior puede considerarse aproximadamente como una $N(0,1)$. Si la muestra es pequeña, no.

5.5.3 Distribución de la varianza muestral

La distribución muestral del estadístico S^2 tiene pocas aplicaciones prácticas en estadística, sin embargo, una ligera modificación resulta de gran utilidad.

Teorema *Teorema de Fisher.*

Sea $(X_1, \dots, X_n)$ una muestra aleatoria simple de tamaño n, procedente de una población $N(\mu,\sigma)$. Entonces se verifica que:

1. Los estadísticos $\bar{X}$ y S^2 son independientes.
2. El estadístico:
$$\frac{(n-1)\cdot S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} \rightarrow \chi^2_{n-1}$$
Sigue una distribución χ^2 con $n-1$ grados de libertad.
3. El estadístico:
$$\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \rightarrow t_{n-1}$$
Sigue una distribución *t*-Student con $n-1$ grados de libertad.

ELENA GONZALO
671 227 871

5.6 DISTRIBUCIÓN DE LA PROPORCIÓN MUESTRAL

Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de tamaño n, procedente de una $B(1,p)$, y sabemos que el estadístico *proporción muestral* será también una variable aleatoria,

$$\hat{p} = P_X = \frac{X}{n}$$

Donde X es el número de elementos de la muestra que poseen la característica que estamos investigando, X sigue una distribución binomial $B(n, p)$.

Cuando n es grande ($n \geq 30$), resulta que el estadístico proporción muestral sigue una distribución asintóticamente normal, es decir:

$$\hat{p} = P_X = \frac{X}{n} \rightarrow N\left(p, \sqrt{\frac{p \cdot q}{n}}\right)$$

También se verifica, para muestras grandes, que:

$$Z = \frac{\hat{P}_X - p}{\sqrt{\frac{p \cdot q}{n}}} \rightarrow N(0,1)$$

Siendo el error estándar del estadístico proporción muestral,

$$\sigma(\hat{P}) = \sqrt{\frac{p \cdot q}{n}}$$

Que disminuye cuando el tamaño de la muestra aumenta.