

Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

Marzo 2021



6.- Teoría del muestreo

6.- Teoría del
muestreo

Población, muestra
y Muestreo

El mundo de la
Población

El mundo de la
muestra

Criterio de selección
de la muestra

El mundo del
Muestreo

Distribución de
los estadísticos

La estadística tiene tres mundos claramente diferenciados:

- ▶ El mundo de la **Población**.
- ▶ El mundo de la **Muestra**.
- ▶ El mundo del **Muestreo**.¹

Debemos saber en todo momento en qué mundo nos encontramos.

Una población es la totalidad de observaciones individuales sobre las cuales se hacen inferencias y que existen en cualquier parte del mundo o al menos dentro del área de muestreo claramente delimitada en el espacio y en el tiempo.

Las poblaciones pueden ser:

- ▶ Finitas, con un número determinado de elementos.
- ▶ Infinitas, cuyo número de elementos no se puede determinar.

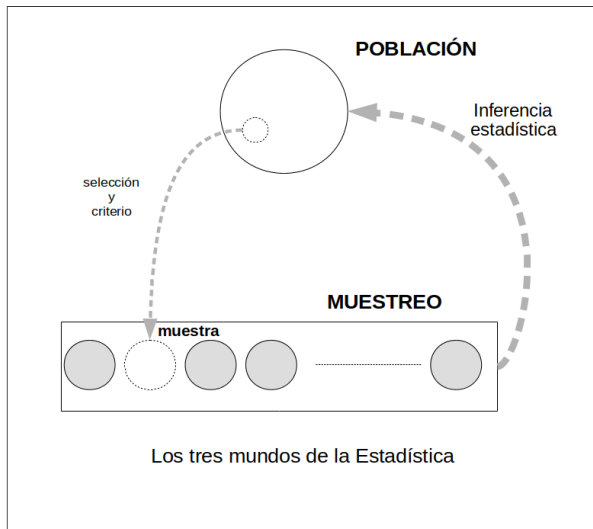
De la población extraemos **muestras**, que son conjuntos de observaciones individuales seleccionadas por un procedimiento específico. Las observaciones individuales miden el **carácter** o **variable** en los individuos de la población. En un mismo individuo, o unidad mínima de muestreo, pueden medirse multitud de variables o caracteres.

“Existe la certeza, que al inferir de la muestra a la población cometemos errores; o dicho de otro modo, la única certeza que tenemos es la de que cometemos errores”².

Pero lo que hay que hacer es controlar el error, no hay que tener miedo al error, sólo hay que controlarlo. Lo que significa que hay que controlar la precisión de la inferencia, controlando todo el **proceso** para tener una determinada confianza o credibilidad en la inferencia al hacer la “extrapolación”.

.^{E1} objetivo de la estadística es conocer lo desconocido,
que es la población".

Los **tres mundos** de la Estadística son:



Población, muestra y Muestreo

6.- Teoría del
muestreo

Población, muestra
y Muestreo

El mundo de la
Población

El mundo de la
muestra

Criterio de selección
de la muestra

El mundo del
Muestreo

Distribución de
los estadísticos

El mundo de la Población

Normalmente el mundo de la población nos es desconocido, mientras que la muestra siempre es conocida (**Estadística descriptiva**) y por eso necesitamos extraer muestras y tener una teoría que nos permita “conocer” a la población (**Estadística inferencial**).

Las poblaciones pueden ser:

1. Finitas: con número finito de elementos N , que es el tamaño de la población³.
2. Infinitas: con número infinito de elementos, se denominan poblaciones infinitas.

6.- Teoría del
muestreo

Población, muestra
y Muestreo

El mundo de la
Población

El mundo de la
muestra

Criterio de selección
de la muestra

El mundo del
Muestreo

Distribución de
los estadísticos

La población consta de:

1. Elementos.
2. Características de los elementos (X).

Si hago una lista de los elementos (X_i), lo que hago es una distribución de frecuencias. Si son muchos los elementos, hacemos: gráficos, tablas o histogramas⁴ que son diferentes maneras visuales de representar a la población. Así es como empezamos a “resumir” a la población.

Otras maneras de “resumir” los elementos es por medio de los parámetros poblacionales, expresados con letras griegas, como por ej.: media de la población, como μ . Si no, usamos siempre letras en mayúsculas. Otro parámetro típico es la varianza σ^2 . También podemos escribir la varianza (para esquemas de muestreo) como S^2 o varianza modificada.

6.- Teoría del
muestreo

Población, muestra
y Muestreo

El mundo de la
Población

El mundo de la
muestra

Criterio de selección
de la muestra

El mundo del
Muestreo

Distribución de
los estadísticos

$$\mu = \frac{\sum X_i}{N} \quad \mu \text{ (Media aritmética)} \quad (1)$$

$$\sigma^2 = \frac{\sum (X_i - \mu)^2}{N} \quad \sigma^2 \text{ (Varianza)} \quad (2)$$

$$\sigma = \sqrt{\sigma^2} \quad \sigma \text{ (Desviación típica)} \quad (3)$$

$$S^2 = \frac{\sum (X_i - \mu)^2}{N - 1} = \frac{S_{XX}}{N - 1} \quad S^2 \text{ (Varianza modificada)} \quad (4)$$

$$S = \sqrt{S^2} \quad (5)$$

$$CV = \frac{\sigma}{\mu} \quad CV \text{ (Coeficiente de variación)} \quad (6)$$

$$CV^2 = \frac{\sigma^2}{\mu^2} \quad CV^2 \text{ (Variación relativa)} \quad (7)$$

El mundo de la muestra

Estamos en el mundo de la muestra, de tamaño n . Se puede hacer una distribución de los elementos, gráficos etc., o resumirla con estadísticos:

6.- Teoría del
muestreo

Población, muestra
y Muestreo

El mundo de la
Población

El mundo de la
muestra

Criterio de selección
de la muestra

El mundo del
Muestreo

Distribución de
los estadísticos

$$\bar{x} = \frac{\sum x_i}{n} \quad \bar{x} \text{ (Media aritmética muestral)} \quad (8)$$

$$s_{\bar{x}}^2 = \frac{S_{xx}}{n-1} = \frac{\sum (x_i - \bar{x})^2}{n-1} \quad s^2 \text{ (Varianza muestral)} \quad (9)$$

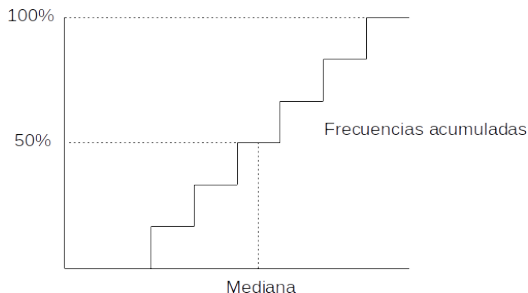
$$s_{\bar{x}} = \sqrt{s_{\bar{x}}^2} \quad s \text{ (Desviación típica muestral)} \quad (10)$$

$$cv_{\bar{x}} = \frac{s_{\bar{x}}}{\bar{x}} \quad cv_{\bar{x}} \text{ (Coeficiente de variación)} \quad (11)$$

$$cv_{\bar{x}}^2 = \frac{s_{\bar{x}}^2}{\bar{x}^2} \quad cv_{\bar{x}}^2 \text{ (Variación relativa. Cochran)} \quad (12)$$

Las ecuaciones (11 y 12) sí que nos permiten **comparar** dispersión entre dos distribuciones.

Cuantil de orden p (cuartiles, quintiles, deciles, percentiles etc.). Si p es 25 % \Rightarrow el cuantil es de orden 25 %, será aquel valor que tenga el 25 % de las frecuencias acumuladas. Por ej. la Mediana es el cuantil de orden 50 %.



Nota sobre la Suma de cuadrados (S_{xx}):

$$S_{xx} = \sum (x_i - \bar{x})^2 \quad \text{definición} \quad (13)$$

$$S_{xx} = \sum x_i(x_i - \bar{x})^2 \quad \text{también teórico} \quad (14)$$

$$S_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n} \quad (15)$$

$$S_{xx} = \sum x_i^2 - n\bar{x}^2 \quad (16)$$

$$S_{xx} = \sum x_i^2 - \bar{x}(\sum x_i) \quad (17)$$

Criterio de selección de la muestra

Debemos de tener siempre un criterio y no extraer la muestra de cualquier manera (criterio probabilístico). Si sacamos la muestra sin criterio, se dice que se saca una muestra al azar⁵.

Un criterio puede ser, el criterio aleatorio simple: donde cualquier elemento de la población tiene la misma probabilidad de ser seleccionado y formar parte de la muestra. Se dice entonces, que se extrae una muestra aleatoria simple (m.a.s. o a.s.).

Otros tipos de criterios de selección de la muestra pueden ser: estratificado, por conglomerados, multietápico o polietápico, sistemático, proporcional etc.

Cada elemento extraído de la población tiene una probabilidad conocida, si esta probabilidad es igual para todos los elementos, decimos que es un m.a.s. y si otorgamos diferente probabilidad por ej.: para las filas de clase, tenemos que la 1ª fila tiene el 10 %, la 2ª fila el 5 %, etc. sería una muestra aleatorio pero no será simple.

Para tener un criterio, la muestra tiene que ser siempre aleatoria, es decir: tener una probabilidad conocida. Si ésta es igual para todos los elementos es una m.a.s. o también, cuando todas las muestras de tamaño n tienen la misma probabilidad de ser seleccionadas.

Existen dos maneras de extraer las muestras:

- ▶ Muestras con reposición
- ▶ Muestras sin reposición

Por ejemplo, tenemos una Población P con tamaño $N = 5$, cuyas puntuaciones son: $X_i = \{5, 8, 2, 7, 3\}$ y extraemos una muestra de tamaño $n = 3$ como puede ser el subconjunto $(5, 8, 2)$. En total, tendremos 10 muestras posibles⁶ $\Rightarrow (5, 8, 2) \dots (2, 7, 3)$

El criterio elegido es un a.s., puede ser expresado como: que cualquiera de las 10 muestras posibles tiene la misma probabilidad de ser sacada. Es decir, que todas las muestras posibles tienen la misma probabilidad; pero para saber esto, tengo que conocer cómo es la población, lo que habitualmente no suele suceder.

El proceso anterior es un ejemplo de a.s. **sin** reposición. Otro sistema diferente es **con** reposición, donde cada elemento puede aparecer más de una vez en la muestra, en cambio sin reposición, cada elemento sólo aparece una vez.

Muestra a.s. posibles	Con reposición	$C'_{m,n} = m^n = N^n$
	Sin reposición	$C_{m,n} = \frac{m!}{(m-n)!n!} = \binom{N}{n}$

Teóricamente es más fácil estudiar la muestra con reposición que sin reposición, pero esto tiene importancia solamente cuando la diferencia entre población y muestra es pequeña. Si estas diferencias son grandes, no existe prácticamente diferencia entre con o sin reposición.

Al cociente $\frac{n}{N}$ se le denomina fracción de muestreo y a $(1 - \frac{n}{N})$ se le llama fracción de corrección para poblaciones finitas.

El mundo del Muestreo

6.- Teoría del
muestreo

Población, muestra
y Muestreo

El mundo de la
Población

El mundo de la
muestra

Criterio de selección
de la muestra

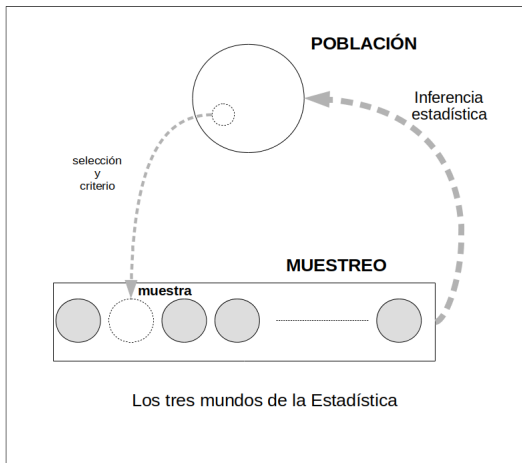
El mundo del
Muestreo

Distribución de
los estadísticos

El **Muestreo** es el conjunto de todas las muestras posibles.

Por ejemplo, de las 10 muestras anteriores, el conjunto de esas 10 muestras posibles es el **Muestreo**. Vamos a introducirnos en la teoría del muestreo que nos enseñará muchas cosas interesantes.

Decíamos que los **tres mundos** de la Estadística eran:



Que a su vez son los tres mundos de/para la inferencia estadística.

Lo que conozcamos de la teoría del muestreo(TM), del mundo del muestreo, será lo que nos servirá para hacer **inferencia estadística**. La inferencia se hace fundamentalmente en el mundo del muestreo, aunque se haga (“aparentemente”) a partir de la muestra. Se hace “aparentemente desde” la muestra porque conozco el muestreo (el conjunto de las muestras posibles).

En la mano la muestra,
pero la mente en el mundo del MUESTREO

Llamamos estadísticos a la: \bar{x} , s^2 , s , *mediana* etc. pero en general los llamaremos **u**. Extraigo de una muestra el estadístico u , pero podría haber sacado otras muestras posibles también con su estadístico u_i , ya que cada muestra del Muestreo puede tomar un valor del estadístico u_i .

6.- Teoría del muestreo

Población, muestra
y Muestreo

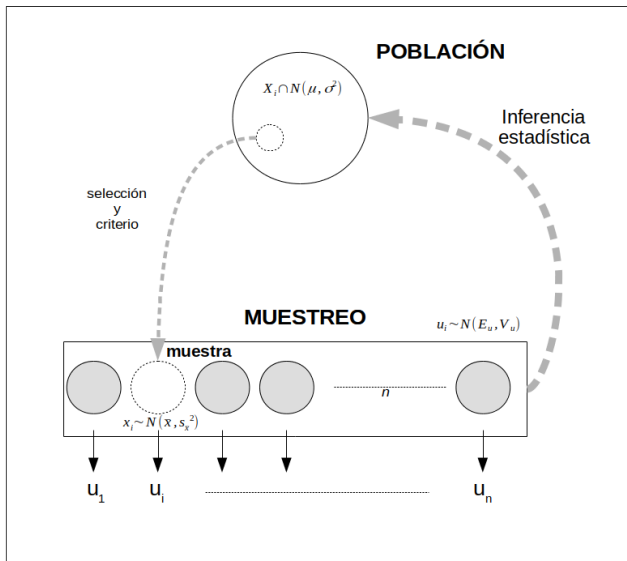
El mundo de la
Población

El mundo de la
muestra

Criterio de selección
de la muestra

El mundo del
Muestreo

Distribución de
los estadísticos



Pero sólo conozco un u (el que extraigo), pero si tuviera todos los estadísticos u_i que puedo obtener de cada una de las muestras posibles, podría hacer una lista, una gráfica o un histograma, esto es: tener y representar la distribución del estadístico u en el muestreo. Es decir, tener una la lista de todos los valores de u de todas las muestras posibles.

Esa distribución de u es fundamental, para la inferencia estadística. En esencia esto es de lo que nos habla la **teoría del muestreo**, que a su vez habla el lenguaje de la probabilidad, la base fundamental de la estadística.

En resumidas cuentas, llamamos a:

- ▶ Las características en la población => **parámetros**
- ▶ Las características en la muestra => **estadísticos**
- ▶ Las características en el muestreo => **valores esperados**
o parámetros (en el muestreo)

Distribución de los estadísticos

Para estudiar la distribución del estadístico u en el **MUESTREO**, recorro a la teoría de probabilidad que nos da la distribución de u en el Muestreo, es decir la Teoría del Muestreo (TM) nos responde sobre cómo es u :

6.- Teoría del
muestreo

Población, muestra
y Muestreo

El mundo de la
Población

El mundo de la
muestra

Criterio de selección
de la muestra

El mundo del
Muestreo

Distribución de
los estadísticos

► La **Esperanza**

- La media de la distribución del estadístico u en el Muestreo μ_u o el más usado $E[u]$, que es el **valor esperado o esperanza del estadístico u** en el Muestreo (la media de la distribución del estadístico u).

► La **Varianza**

- Para la varianza: $\sigma_u^2 = V[u]$, **varianza de la distribución del estadístico u** en el Muestreo.

► El **Error**

- La "desviación típica.^{en} el Muestreo: $\sigma_u = \sqrt{\sigma_u^2}$ se llama **error del estadístico u** en el Muestreo que normalmente se representa como: $\mathcal{E}_u = \sqrt{V[u]}$, a veces se denomina simplemente error de u , o sea error es el nombre que damos a la desviación típica de la distribución de u en el Muestreo.

► El **Comportamiento**

- En el Muestreo podemos hacer afirmaciones como cuales: $Prob\{u \leq ?\} = 0,99$. Es decir, estamos hablando de que sabemos **cómo se distribuye u** .

Conclusión:

**¡Hay que separar claramente los tres
mundos de la Estadística!**

¡Nunca debemos confundirlos!



6.- Teoría del muestreo

Población, muestra
y Muestreo

El mundo de la
Población

El mundo de la
muestra

Criterio de selección
de la muestra

El mundo del
Muestreo

Distribución de
los estadísticos