



# **INTRODUCCIÓN AL MUESTREO Y ESTIMACIÓN PUNTUAL E INTERVALAR**

*Penna – Cobos – Vázquez Ferrero – Ulagnero*



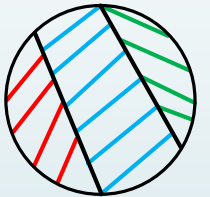
- ☐ **Muestreo aleatorio, probabilístico o al azar:** cuando todas las UA que componen la población tienen igual probabilidad de pertenecer a la muestra.
- ☐ **Muestreo no aleatorio, no probabilístico o no al azar:** cuando no todas las UA que componen la población tienen igual probabilidad de pertenecer a la muestra.

# MUESTREO PROBABILÍSTICO

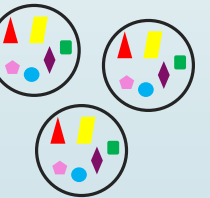
➤ **Muestreo aleatorio simple:** consiste en seleccionar “n” UA de las “N” que conforman la población de manera que todas ellas tengan igual probabilidad de pertenecer a la muestra.



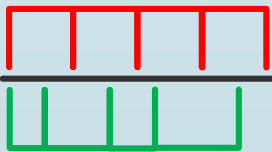
➤ **Muestreo aleatorio estratificado:** consiste en dividir la población en subpoblaciones o estratos, tal que cada estrato sea internamente homogéneo, mientras que los estratos sean heterogéneos entre sí. Luego se realiza un muestreo aleatorio simple en cada conglomerado.



➤ **Muestreo por conglomerados:** consiste en dividir la población en subpoblaciones llamadas conglomerados, cuya característica es ser internamente heterogéneo y homogéneos entre sí.

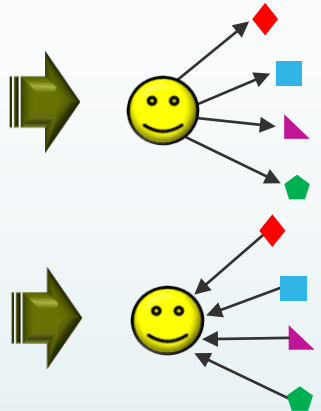


➤ **Muestreo sistemático:** consiste en seleccionar aleatoriamente la primer UA para la muestra, y luego seleccionar las UA posteriores utilizando intervalos fijos o aleatorios hasta llegar el tamaño de muestra deseado.



# MUESTREO NO PROBABILÍSTICO

- **Muestreo intencional:** se da cuando el investigador escoge aquellas UA que considera típicas del grupo a estudiar.
- **Muestreo accidental:** se da cuando el investigador toma las UA que estén disponibles en el momento.



## ALGUNAS CONSIDERACIONES

Ningún tipo de muestreo es mejor que el otro y su utilización dependerá de varios factores, entre ellos las características de la población y, en muchos de los casos, la posibilidad de manejar los aspectos técnicos del diseño muestral.

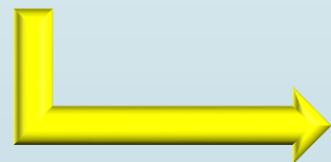
Por ejemplo, si estamos interesados en realizar inferencias, el muestreo deberá ser probabilístico; pero si nuestro interés radica en un estudio exploratorio o descriptivo, podemos utilizar un muestreo no probabilístico.

# DISTRIBUCIONES MUESTRALES

Como ya se dijo, el muestreo (cuando es probabilístico) muchas veces tiene como objetivo inferir propiedades de la población a partir de una muestra. Estadísticamente, se pretende conocer los parámetros de la distribución de la variable de interés y es el muestreo el encargado de proveer dicha información.

Luego, los estadísticos muestrales sirven como estimadores de los parámetros que caracterizan a la distribución. Por otra parte, los estadísticos son variables aleatorias y como tales, tienen una distribución asociada.

Recordemos que:



**Media Aritmética →**

**Varianza →**

**Desviación estándar →**

**Mediana →**

...

**Proporción →**

Estimadores (n)

$\bar{X}$

$S^2$

$S$

$\tilde{X}$

...

$\hat{p}$

Parámetros (N)

$\mu$

$\sigma^2$

$\sigma$

$\Theta$

...

$\Pi$

# DISTRIBUCIÓN DEL ESTADÍSTICO MEDIA MUESTRAL ( $\bar{X}$ )

Ya que  $\bar{X}$  es una variable aleatoria, nos interesa conocer su distribución. Cuando se estudian las distribuciones de los estadísticos muestrales se hace desde un punto de vista teórico, suponiendo poblaciones infinitas. Si se quieren observar estas propiedades partiendo de poblaciones finitas, se recurre a un muestreo con reposición porque de esa forma se emula una población de tamaño infinito.

## DEFINICIÓN DE ERROR ESTÁNDAR

La desviación estándar de las medias de la muestra de tamaño  $n$ , recibe el nombre de Error Estándar y se define como:

$$EE = \sqrt{\sigma_{\bar{X}}^2} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

Directamente proporcional a la desviación estándar e inversamente proporcional al tamaño de la muestra de la cual se calcula

# DISTRIBUCIÓN DEL ESTADÍSTICO MEDIA MUESTRAL ( $\bar{X}$ )

## TEOREMA CENTRAL DEL LÍMITE (TCL)

Sea  $X$  una variable aleatoria con  $E(X) = \mu$  y  $V(X) = \sigma^2$ . Sea  $\bar{X}$  la media de una muestra aleatoria de tamaño  $n \Rightarrow$  la variable aleatoria,  $Z$ , se define como:

$$Z = \left( \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \right) \xrightarrow{n \rightarrow \infty} N(0, 1)$$

**Nota:** lo interesante de éste teorema es que no hace referencia a la distribución de  $X$ , es decir que aunque  $X$  no se distribuya normalmente, si tiene varianza finita entonces, para  $n$  suficientemente grande,  $Z$  converge a una  $N(0,1)$ .

# DISTRIBUCIÓN DEL ESTADÍSTICO MEDIA MUESTRAL ( $\bar{X}$ )

## DISTRIBUCIÓN “t” DE STUDENT

La mayor dificultad en aplicar el resultado anterior es que, en la práctica,  $\sigma^2$  es desconocida. Luego se podría estimar su valor a partir de una muestra, lo cual se logra sustituyendo en la fórmula anterior  $\sigma$  por  $S$ . El problema es que dicha sustitución, modifica la variable aleatoria  $Z$  a la que hace referencia el TCL y por ende ya no se tiene una distribución normal para la estandarización. La variable aleatoria:

$$t = \left( \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \right) \rightarrow t_{n-1}$$

**Nota:** cuando los grados de libertad de una distribución “t”  $> 30$ , la forma de la distribución, se aproxima a la  $N(0; 1)$ .

# EJEMPLO

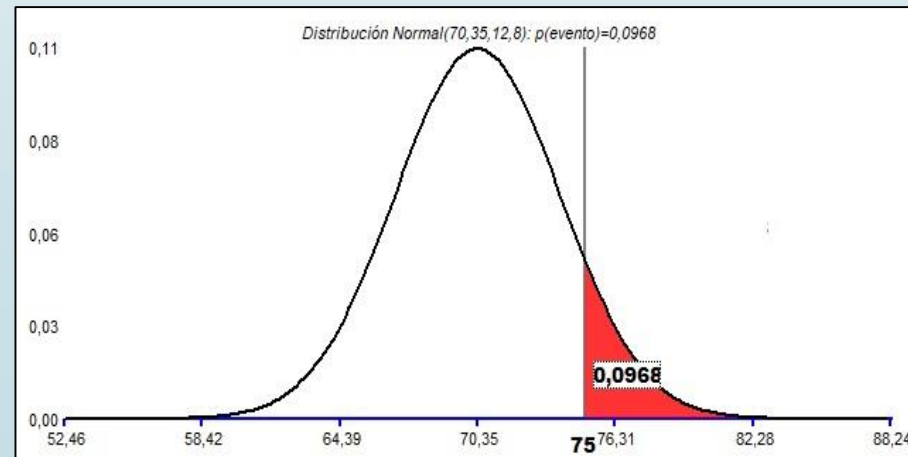
La distribución de la variable aleatoria “nota de una determinada materia” se aproxima a una distribución  $N(70,35; 8)$ .

a) ¿Cuál es la probabilidad que la media, de una muestra de 5 alumnos, exceda el valor 75?

$$P(\bar{X} > 75) = P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} > \frac{75 - \overset{\text{EE}}{70,35}}{\underset{\text{EE}}{8/\sqrt{5}}}\right) \cong P(Z > 1,30) = \Phi(1,30) = \mathbf{0,0968}$$

Tabla 1

Entonces, la probabilidad que la media muestral exceda 75, es de 0,0968.



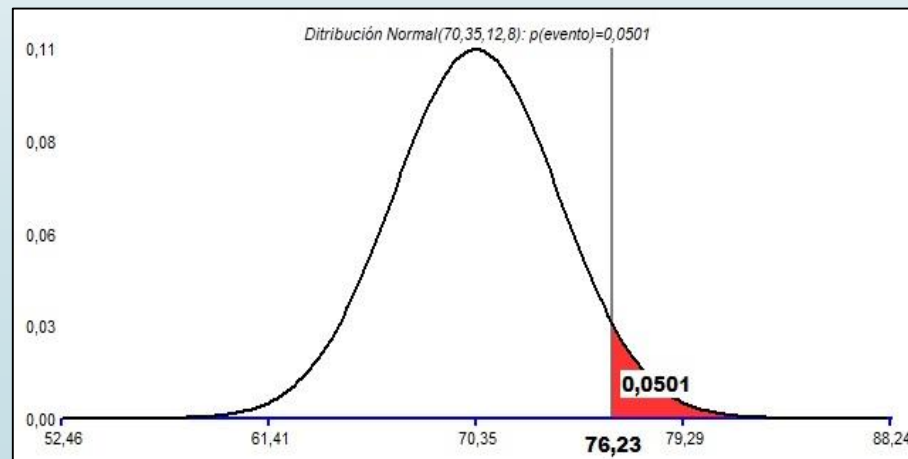
## EJEMPLO (CONT.)

b) ¿Cuál es la nota promedio sólo superada por un 5% de las notas promedio?

**Tabla 1**  $\Rightarrow Z_0 = 1,64 \Rightarrow \frac{\bar{X} - 70,35}{8/\sqrt{5}} = 1,64 \Rightarrow \bar{X} \cong 76,23$

**EE**

La nota promedio, sólo superada por un 5%, es de 76,23.



## EJEMPLO (CONT.)

Supongamos ahora, que no se conoce el valor de la desviación estándar de la población ( $\sigma$ ) pero se dispone de la siguiente muestra para estimarla:

67,9 69,3 70,0 74,8 75,3 69,6 67,3 65,8 70,5

c) ¿Cuál es la probabilidad que la media, de una muestra de 5 alumnos, exceda el valor 75?

La desviación estándar, obtenida de la muestra, es:  $S \cong 3,19 \Rightarrow$

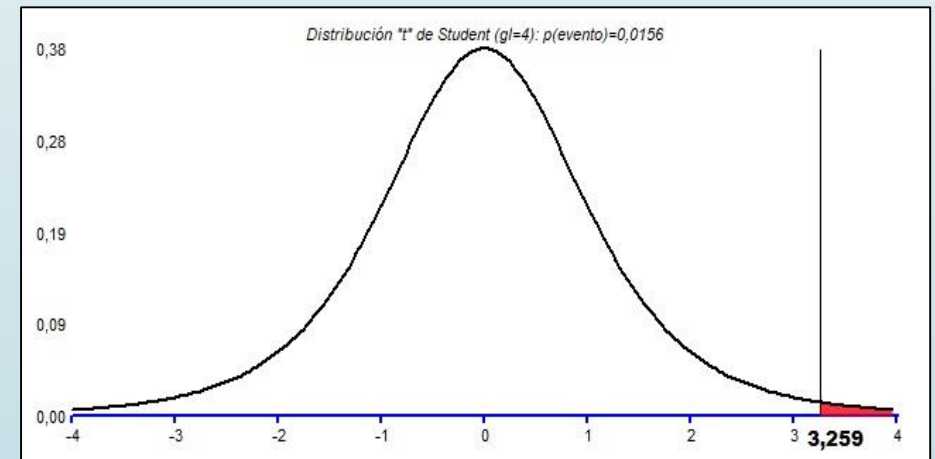
$$P(\bar{X} > 75) = P\left(\frac{\bar{X} - \mu}{S/\sqrt{n}} > \frac{75 - 70,35}{3,19/\sqrt{5}}\right) = P(t_4 > 3,259) \cong 0,01$$

Tabla 2

Entonces, la probabilidad que la media muestral exceda 75, es de 0,01.

EE

InfoStat



## EJEMPLO (CONT.)

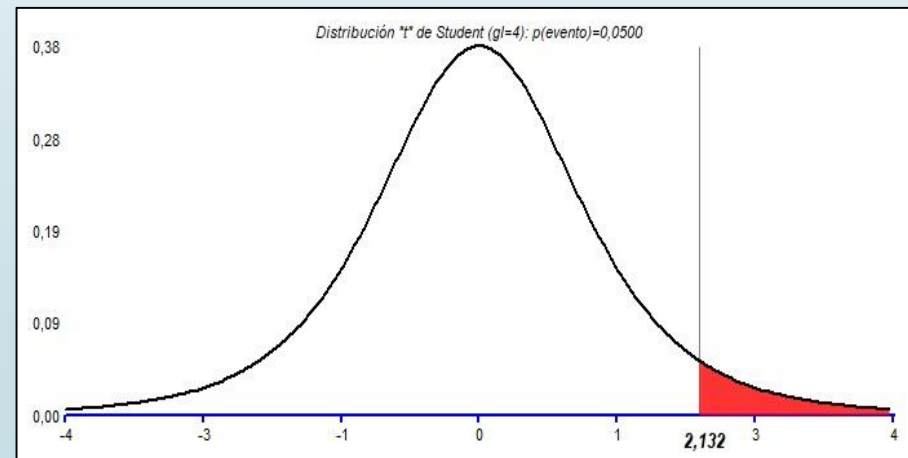
d) ¿Cuál es la nota promedio sólo superada por un 5% de las notas promedio?

Tabla 2

$$P(t_4 > t_0) = 0,05 \Rightarrow t_0 = 2,132 \Rightarrow \frac{\bar{X} - 70,35}{3,19/\sqrt{5}} = 2,132 \Rightarrow \bar{X} \cong 73,39$$

EE

La nota promedio, sólo superada por un 5%, es de 73,39.



# ESTIMACIÓN DE PARÁMETROS

La inferencia estadística se interesa por dar conclusiones de un gran número de acontecimientos, fundándose en las observaciones de una parte de los mismos.

La estadística nos proporciona herramientas que formalizan los procedimientos para sacar conclusiones, siempre que las muestras seleccionadas sean representativas de la población que han sido extraídas. Esta representatividad permite que los valores que describen a las muestras se extiendan a la población correspondiente, es decir que permiten realizar una inferencia.

Los estadísticos, valores obtenidos en la muestra, son estimadores de los parámetros correspondientes. Vamos a centrarnos en dos tipos de estimaciones: *puntual e intervalar*.

## ESTIMACIÓN PUNTUAL

Considera un único valor como estimación del parámetro; es decir se usa un solo estadístico muestral para estimar el parámetro poblacional correspondiente.

# ESTIMACIÓN DE PARÁMETROS

## PROPIEDADES CLÁSICAS DE LOS “BUENOS” ESTIMADORES

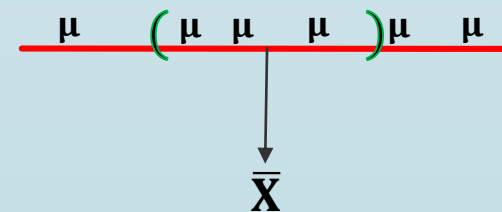
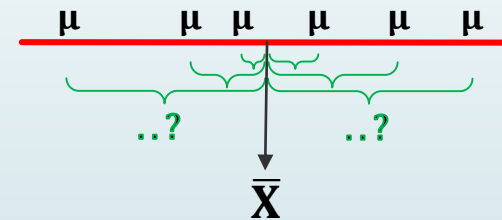
- **Insesgado:** Un estimador es *insesgado* si el promedio, de las medias de todas las muestras posibles de tamaño  $n$  de una población, es igual al parámetro. [InfoStat](#)
- **Consistente:** Un estimador es *consistente* en la medida en que, al aumentar el tamaño de la muestra, su valor se acerca cada vez más al parámetro correspondiente.
- **Eficiente (o de varianza mínima):** Un estimador será más *eficiente* (preciso) cuanto menor sea la variabilidad de muestra a muestra de una misma población. Como la variabilidad de una distribución muestral viene dada por su error estándar, un buen estimador será aquel que menor error estándar alcanza.
- **Suficiente:** Un estimador es *suficiente* cuando contiene la suficiente información de la muestra de la cual fue calculado.

# ESTIMACIÓN DE PARÁMETROS

## ESTIMACIÓN POR INTERVALOS

Hasta ahora hemos ofrecido un solo valor, un solo punto, como estimación del parámetro que se trate. Por otra parte, no podemos indicar la diferencia probable (o distancia) entre el parámetro y el estadístico, es decir el error probable cometido al estimar el parámetro. Lo único que podemos afirmar es que, en la mayoría de los casos, ese error tenderá a disminuir a medida que vaya aumentando el tamaño de la muestra.

En la estimación por intervalos vamos a ofrecer una infinidad de valores, un intervalo de puntos, dentro del cual esperamos se encuentre el parámetro. Además indicaremos la probabilidad o confianza con la que esperamos que el intervalo contenga el valor del parámetro.



# ESTIMACIÓN DE PARÁMETROS

## PROCEDIMIENTO PARA ENCONTRAR UN INTERVALO DE CONFIANZA PARA LA MEDIA POBLACIONAL ( $\mu$ )

El objetivo del procedimiento de estimación por intervalo es encontrar el intervalo cerrado  $[L_I; L_S]$  donde  $L_I$  = Límite Inferior y  $L_S$  = Límite Superior, tal que si el parámetro a estimar es  $\mu$ , entonces:

$$P(L_I \leq \mu \leq L_S) = P(\mu \in [L_I; L_S]) = 1 - \alpha$$

- *Límites de confianza*: son los dos extremos (inferior y superior) del intervalo de confianza.
- *Intervalo de confianza*: es el intervalo dentro del cual confiamos contenga al parámetro que va a ser estimado.
- *Nivel de confianza* ( $1 - \alpha$ ): es la probabilidad o grado de confianza según el cual afirmaremos que el parámetro se encuentra dentro del intervalo.

## REFERENCIAS BIBLIOGRÁFICAS

- Bologna, E. (2011). *Estadística para Psicología y Educación*. Córdoba: Brujas.
- Gorgas García, J., Cardiel López, N. & Zamorano Calvo, J. (2009). *Estadística Básica para Estudiantes de Ciencias*. Madrid: Departamento de Astrofísica y Ciencias de la Atmósfera. Facultad de Ciencias Físicas. Universidad Complutense de Madrid.
- Penna, F.O., Esteva, G.C., Cobos, O.H. & Ulagnero, C.A. (2018). *Fórmulas y Tablas III (para cursos de Estadística básica)* (2ª ed.). San Luis: Nueva Editorial Universitaria.
- Sabulsky, J. (2000) *Investigación científica en salud-enfermedad*. (3ª ed.). Córdoba: Ed. Kosmos.
- Triola M. (2018). *Estadística* (12ª ed.). México: Pearson Educación.