# LA INFERENCIA ESTADÍSTICA Y LAS TIC (1ª PARTE)

ROSANA ÁLVAREZ GARCÍA (*) y ABEL MARTÍN (**)

## CONCEPTOS BÁSICOS

Para comenzar a trabajar la Inferencia Estadística iniciaremos con el alumnado un repaso de conceptos básicos que utilizaremos frecuentemente durante el desarrollo del tema:

- **Población:** conjunto de todos y cada uno de los elementos que nos interesa estudiar y que poseen una o más características comunes.
- **Individuos, unidades, elementos:** cada uno de los elementos que constituyen la población.
- **Tamaño de la población (N):** es el número total de individuos que constituyen dicha población.
- **Carácter:** características y cualidades de los individuos de la población que queremos estudiar. Cada carácter puede tomar distintos valores o modalidades. Una variable estadística recorre todos los valores de un cierto carácter.
- **Tipos** de variables ESTADÍSTICAS.
  - **Variable estadística cuantitativa:** Toma valores numéricos.
    - **(a)** V. e. c. discreta: Toma valores numéricos concretos.
    - **(b)** V. e. c. continua: Pueden tomar cualquier valor en un intervalo.
  - **Variable estadística cualitativa:** No toma valores numéricos.
- **Muestra:** en muchas ocasiones es muy difícil estudiar todos y cada uno de los componentes de una población, por eso tomamos muestras, subconjuntos de la población en estudio.

## MUESTREO

La **Estadística Descriptiva** se encarga de la recogida, organización y análisis de los datos obtenidos de la población en estudio y la Inferencia Estadística de extraer conclusiones a partir de los datos obtenidos.

(*) Profesora de Tecnología del IES Aramo (Oviedo-Asturias).

(**) Profesor de Matemáticas del IES Pérez de Ayala (Oviedo-Asturias).

Dentro de la Inferencia Estadística tenemos:

- La estimación de parámetros poblacionales:

  1. Estimación puntual.
  2. Estimación por intervalos.

- Contraste o test de hipótesis, procedimiento que nos permite rechazar o no rechazar una hipótesis sobre algún parámetro poblacional a partir de los datos obtenidos en muestras extraídas de la población.

  Para llevar a cabo la recogida de datos de una población trabajamos con muestras debidamente elegidas. Las razones por las que se trabaja con muestras poblacionales son:

- La imposibilidad de trabajar con todos los elementos de la población por su tamaño, coste económico del estudio, tiempo necesario para su realización, etc.
- Las poblaciones son homogéneas, lo que permite obtener buenos resultados con las muestras extraídas.

Entre los inconvenientes que conlleva el trabajo con muestras y no con el conjunto de la población podríamos destacar:

- La importancia de que las muestras sean representativas de la población para que los datos obtenidos permitan realizar un análisis correcto y poder extrapolar los resultados a la población sin cometer errores.
- La necesidad de conocer y dominar las diferentes técnicas de muestreo a utilizar en cada tipo de estudio según las variables que se van a estudiar y el tipo de población con el que vamos a trabajar, con el fin de minimizar los errores en el estudio y en las posteriores conclusiones.

## TIPOS DE MUESTREO MÁS UTILIZADOS

### Muestreo con reemplazamiento

Cada elemento o individuo observado se reintegra a la población, pudiendo formar parte nuevamente de la muestra. En poblaciones pequeñas este método puede presentar graves inconvenientes como, por ejemplo, que en una encuesta le preguntáramos varias veces a la misma persona por el tema de interés.

### Muestreo sin reemplazamiento

Cada individuo observado es forzosamente distinto de los demás. Si el tamaño de la población es muy grande no presenta grandes diferencias con el muestreo aleatorio con reposición porque la probabilidad de repetir el individuo es muy pequeña. Matemáticamente, ambos modelos coinciden en poblaciones infinitas que son las que constituyen el objeto de nuestro estudio.

## Muestreo aleatorio simple

Todos los miembros de la población tienen la misma probabilidad de aparecer en la muestra. Un ejemplo claro lo vemos en los muestreos realizados por Internet: no todo el mundo dispone de una conexión a Internet, incumpliéndose el principio básico de este tipo de muestreo (todos los elementos de la población tienen la misma probabilidad de formar parte de la muestra elegida).

Dentro del muestreo aleatorio simple tenemos:

- **Muestro por cuotas.** Se fija de antemano el número de individuos a muestrear dentro de los que cumplen ciertos requisitos; presenta el inconveniente de que la determinación de tales requisitos quede al arbitrio del seleccionador de la muestra, incumpliéndose así los planteamientos objeto de los modelos teóricos y, por tanto, cuestionándose la validez de sus conclusiones.
- **Muestreo estratificado.** La población se divide en grupos más pequeños o estratos homogéneos. Una vez establecidos los diferentes estratos se obtiene una muestra aleatoria simple de cada uno de ellos. La muestra total obtenida es la suma de las muestras extraídas de cada estrato. Un claro ejemplo lo constituyen los llamados estratos sociales muy empleados en encuestas de opinión (por edades, por sexos, por niveles de estudios, etc.).

  Dentro de este tipo de muestreo podemos diferenciar a su vez:

  - **Muestreo aleatorio estratificado constante.** Las muestras extraídas de cada estrato tienen el mismo tamaño.

    $$n_1 = n_2 = \ldots = n_z = n/Z$$

  - **Muestreo aleatorio estratificado proporcional.** Las muestras extraídas de cada estrato son proporcionales a su tamaño.

    En el caso de que haya "k" estratos y que el número de elementos de cada estrato sea $N_1$ , $N_2$ , ... $N_k$ ; si n es el tamaño de la muestra y N el tamaño de la población, se calculará el número de elementos que tomaremos en cada muestra de forma proporcional al tamaño de los estratos (**afijación proporcional**) con la ayuda de la expresión:

    $$n_i = n \cdot \frac{n_i}{N}$$

- **Muestreo sistemático.** Se selecciona al azar un elemento de la población y a partir de este elemento se seleccionan de k en k. En este caso sólo es necesario seleccionar un elemento al azar de la muestra.
- **Muestreo por conglomerados.** Se divide la población en unidades heterogéneas como la propia población. Es frecuente que los conglomerados vengan dados por las divisiones administrativas (Comunidades Autónomas, Ayuntamientos, etc.), o por razones geográficas claramente definidas, de ahí que también se le llame muestreo por áreas. El reparto se hace generalmente de forma proporcional a los tamaños de los conglomerados y, en cada uno de ellos, se muestrea aleatoriamente.
- **Muestreo bietápico.** Es, en cierto modo, una variante del anterior, en la que no se muestrea en todos los conglomerados, sino en una parte de ellos seleccionados aleatoriamente. Cuando cada conglomerado se puede considerar formado por otro conglomerados de menor entidad, se puede ampliar el número de etapas para dar lugar a un muestreo polietápico.

### Muestreo no aleatorio

En la muestra seleccionada no todos los elementos tienen la misma probabilidad de ser incluidos. La representatividad en este tipo de muestreo es escasa y las muestras obtenidas no son muy válidas para realizar inferencias. Dentro de este tipo de muestreo tenemos:

- **Muestreo intencional u opinático.** El investigador selecciona la muestra que supone es la más representativa, utilizando un criterio subjetivo y en función de la investigación que se vaya a realizar.
- **Muestreo sin norma, circunstancial o errático.** Estas muestras se emplean mucho en la vida corriente; por ejemplo, en el comercio, cuando se supone que un trozo de tela o un sorbo de vino representan bien los artículos completos.

## ACTIVIDADES DE MUESTREO

### Actividad Propuesta 1

Ante un próximo referéndum en una población de 120.000 personas, de las que sólo pueden votar 80.000, se desea efectuar un sondeo mediante una encuesta a 40 personas:

(a) Comentar, brevemente, cómo podrá seleccionarse una muestra aleatoria para llevar a cabo dicha encuesta. ¿La selección se hará en toda la población o sólo en la de posibles votantes?. ¿Deberá efectuarse un muestreo con o sin reposición?. ¿Por qué?

(b) Si de las 40 personas encuestadas 25 han respondido que votarán "SI", 10 que votarán "NO" y 5 han manifestado su intención de abstenerse, ¿cuál será la estimación obtenida para el porcentaje de votos afirmativos?, ¿y del total de votos afirmativos?

**Propuesta de resolución**

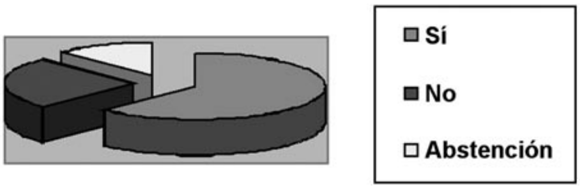
(a) Una forma posible de seleccionar la muestra podría ser la siguiente:

- Numerar a los individuos del censo de votantes (del 1 al 80.000) y generar una lista de 40 números aleatorios utilizando cualesquiera de los procedimientos existentes (tabla de números aleatorios, ordenador, función RANDOM de la calculadora que explicaremos más adelante).
- La muestra estaría formada por los individuos del censo de votantes cuyo número corresponda con alguno de la lista anterior.
- La selección debe hacerse tomando como población a los posibles votantes, ya que es la "población" en estudio.
- El muestreo aleatorio podrá realizarse indistintamente, con o sin reposición, ya que el tamaño de la muestra es suficientemente pequeño con respecto al tamaño de la población y la probabilidad de elegir a un individuo dos veces es casi nula.

(b) Si denotamos por "p" a la proporción de votos afirmativos en la población, su estimación será:

$$p = \frac{n^o \textit{ de votos afirmativos en la muestra}}{\textit{tamaño de la muestra}} = \frac{25}{40} = 0{,}625$$

$$0{,}625 \cdot 80.000 = 50.000$$




Por lo tanto, la estimación puntual del porcentaje de votos afirmativos será del 62.5% y la estimación del total de votos afirmativos de 50 000.

## Actividad propuesta 2

Cierto museo ha organizado una exposición sobre la obra de un pintor contemporáneo. Al objeto de poder evaluar dicha muestra con posterioridad, se pide a cada visitante que rellene un pequeño cuestionario:

(a) En lugar de examinar todas las respuestas recibidas una vez clausurada la exposición, para obtener conclusiones con mayor rapidez, se quiere analizar sólo 60: indicar un método de selección adecuado.

(b) La dirección del museo sospecha que el interés de la exposición es distinto según la edad del visitante. Se han clasificado los cuestionarios por tramos de edad, con los siguientes resultados:

| Edad | Número |
|---|---|
| Menores de 25 | 200 |
| 25–40 | 1.000 |
| 40–60 | 500 |
| Más de 60 | 300 |

(b1) Describir cómo se elegirá la muestra aplicando el muestreo estratificado

(b2) Calcular el tamaño de muestra correspondiente a cada estrato.

**Propuesta de resolución**

(a) Si no tenemos razones para pensar que algún factor pueda influir en la opinión de los encuestados respecto del museo, podemos llevar a cabo un muestreo aleatorio simple. En otro caso, estratificaremos teniendo en cuenta ese factor (edad, formación cultural, sexo, etc.).

(b) Tomaremos como estratos los distintos tramos de edades, y en cada uno de ellos se realizará un muestreo aleatorio simple con *afijación proporcional* al tamaño del estrato.

| EDAD | $N_i$ | $n_i = n \cdot N_i / N$ | Tamaño de muestra $n_i$: |
|---|---|---|---|
| **Menores de 25** | 200 | 60 · 200/2.000 | 6 |
| **25 – 40** | 1.000 | 60 · 1.000/2.000 | 30 |
| **40 – 60** | 500 | 60 · 500/2.000 | 15 |
| **Más de 60** | 300 | 60 · 300/2.000 | 9 |

## Actividad propuesta 3

Se considera un grupo de 240 personas consumidoras de un determinado producto. Queremos, para facilitar el estudio que vamos a realizar, elegir una muestra compuesta por 20 de ellas:

(a) ¿Cómo podríamos hacer un muestreo aleatorio simple usando una tabla de números aleatorios?

(b) ¿Cómo haríamos un muestreo sistemático?

**Propuesta de resolución**

(a) Elegimos para trabajar una tabla de números aleatorios que podemos encontrar en algún libro de Bachillerato o confeccionarla nosotros mismos con una calculadora científica o con una calculadora gráfica, tal y como explicaremos al acabar esta actividad. Escogemos una columna de la tabla y en ella los tres primeros dígitos de un número que sean menores de 241 pues nosotros trabajaremos con 240 personas, procurando que no se repitan; así el estudio sería sin reposición, pudiendo obtener la siguiente tabla:

239, 064, 211, 182, 001, 025, 109, 147, 134, 123, 172, 197, 171, 033, 086, 149, 047, 045, 051, 218

ordenados, los elementos elegidos aleatoriamente serían los siguientes:

1, 25, 33, 45, 47, 51, 64, 109, 123, 134, 147, 149, 171, 172, 182, 197, 211, 218, 239

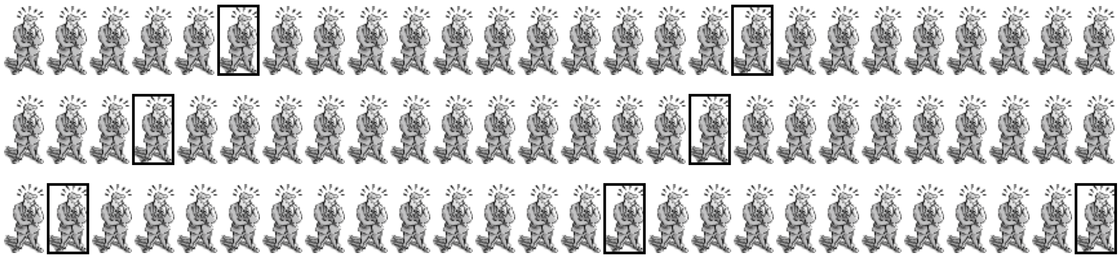
(b) Al trabajar con 240 personas y elegir una muestra de 20, calculamos k:

$$k = \frac{N}{n} = \frac{240}{20} = 12$$

Se elige aleatoriamente un número entre uno y doce, ambos incluidos. Tomamos por ejemplo el número 6. La muestra quedará compuesta del modo siguiente:

6, 18, 30, 42, 54, 66, 78, 90, 102, 114, 126, 138, 150, 162, 174, 186, 198, 210, 222, 234.

Todos ellos formados sumándole 12 al primero elegido al azar.

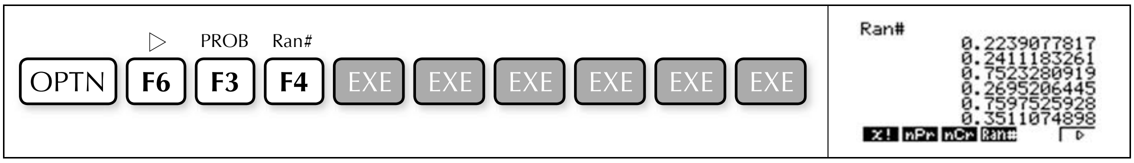


## APRENDIENDO A GENERAR ALEATORIAMENTE NÚMEROS CON UNA CALCULADORA GRÁFICA

¿Sabías que la calculadora es capaz de generar números al azar mediante la función RANDOM?

Iremos poco a poco, pensando y razonando hasta ser capaces de aplicarlo a cualquier situación que se nos plantee.

(1) Buscamos la opción **RAN#** en la calculadora y presionamos sucesivamente **EXE.**



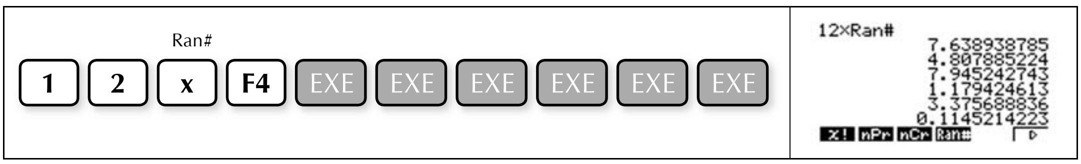

- ¿Podrías dar una característica común a todos estos números que te aparecieron?
- ¿Cuál intuyes que será el número más alto que puede aparecer?
- ¿Y el más bajo?

Me imagino que habrás adivinado que los números que se generan tienen todos 10 cifras decimales y que el número más alto que puede salir es el 0,9999999999 y el número más bajo es el 0,0000000000.

Nos imaginamos que queremos sacar números aleatoriamente para 12 personas.

- ¿Qué podríamos hacer para que el número máximo fuese el 12 y el mínimo el 0?

  Simplemente multiplicando previamente por 12. ¡Compruébalo!

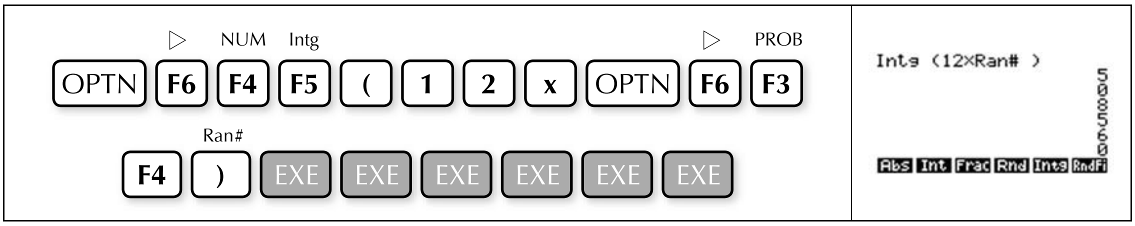



  – Antes el número máximo era 0.9999999999; ahora el número máximo será 12x0.9999999999 = 11.9999999

  – Antes el número mínimo era 0, Ahora el número mínimo será 0 x 12 = 0

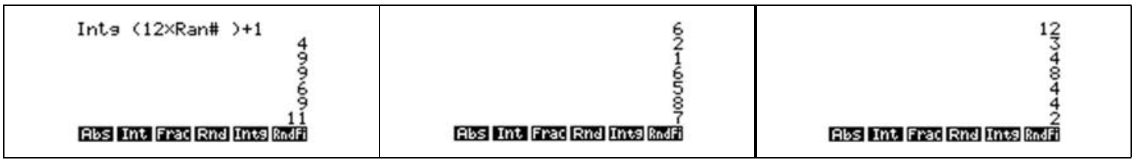
- ¿Cómo conseguimos que estos números no sean decimales?.

Acudimos a la OPCIÓN **Intg** que tiene la calculadora: permite obtener el número entero más grande que no sea mayor al valor.




Pero presenta un inconveniente: aparecerá el 0 y nunca aparecerá el 12. ¿cómo se puede solucionar?

Simplemente sumándole 1 a la expresión:



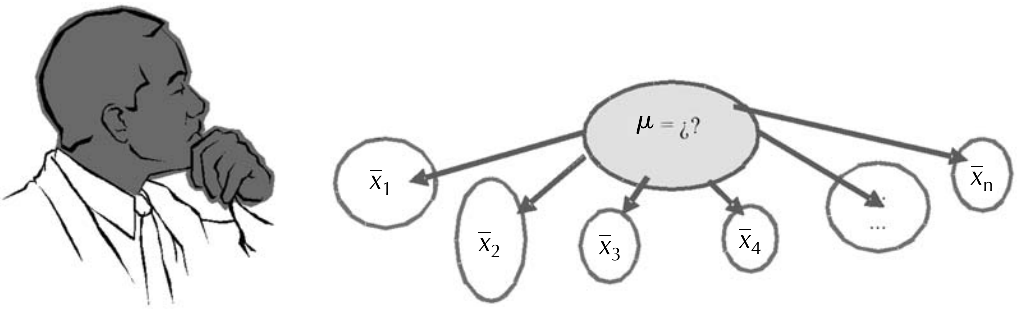

## ESTIMACIÓN DE PARÁMETROS POBLACIONALES

El estudio de las características observadas en una población se lleva a cabo a través de los resultados obtenidos en varias muestras extraídas de ella. Los estadísticos, media aritmética, proporción y desviación típica que obtenemos en las muestras nos permiten obtener aproximaciones de los parámetros poblacionales. Para conseguir la estimación de estos parámetros poblacionales debemos saber su relación con los parámetros muestrales, siendo necesario conocer la distribución muestral de estos estadísticos.

Las distribuciones muestrales de la media y la proporción son.

## DISTRIBUCIÓN MUESTRAL DE LA MEDIA

Extraemos muestras aleatorias de una población cuya media queremos estimar, cada una de las muestras elegidas de tamaño "n" nos proporciona un valor distinto para la media.




Definimos una nueva variable aleatoria simple que toma como valores los valores obtenidos para la media en las muestras extraídas de la población:

$$\overline{X} = \bar{x}_1,\ \bar{x}_2,\ \bar{x}_3,\ \bar{x}_4,\ \ldots,\ \bar{x}_n$$

La media de nuestra variable aleatoria $\overline{X}$ es

$$E\,[\overline{X}] = \sum_{i=1}^{i=n} \frac{\overline{x}_i \cdot n_i}{n}$$

Y la desviación típica

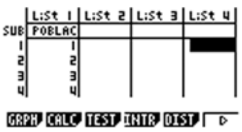
$$\sigma_{\overline{x}} = \sqrt{\sum_{i=1}^{i=n} \frac{\overline{x}_i^{\,2} \cdot n_i}{n} - E\,[\overline{X}]^2}$$

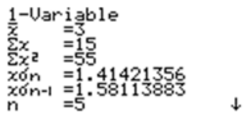
Vamos a observar la relación entre los parámetros muestrales y poblacionales en un ejemplo:

Tenemos una población formada por cinco elementos:

A = { 1, 2, 3, 4, 5 }

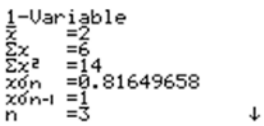
Vamos a calcular la media y desviación para la población:



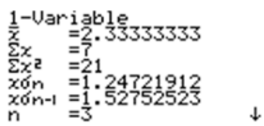




Tomamos todas las posibles muestras de tamaño 3 a partir de la población:

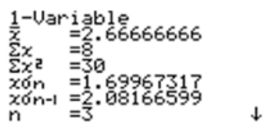
A = {1, 2, 3}



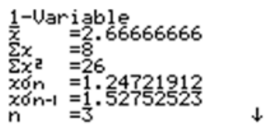

B = {1, 2, 4}



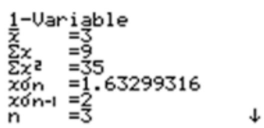

C = {1, 2, 5}



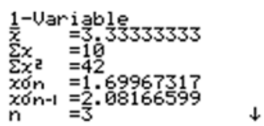

D = {1, 3, 4}



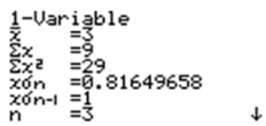

E = {1, 3, 5}



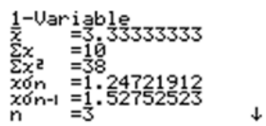

F = {1, 4, 5}



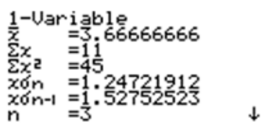

G = {2, 3, 4}



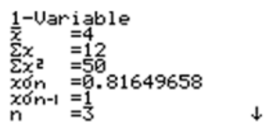

H = {2, 3, 5}




I = {2, 4, 5}



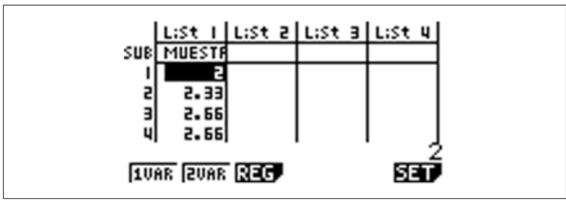

J = {3, 4, 5}

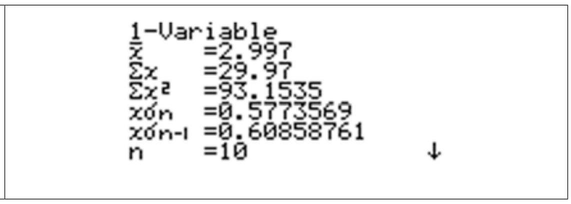



Nuestra nueva variable aleatoria media muestral toma los valores:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $\overline{X}$ | 2 | 2.33 | 2.66 | 2.66 | 3 | 3.33 | 3 | 3.33 | 3.66 | 4 |

La media y desviación típica de nuestra nueva variable es:




$E[\overline{X}] = 2.997 \approx 3$ $\sigma_{\overline{x}}\,[\overline{X}] = 0.5773569$

## TEOREMA CENTRAL DEL LÍMITE

Este teorema nos garantiza que la distribución de $\overline{X}$ (parámetro a estimar) está centrada en el mismo lugar que la variable en estudio $X$.

Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución con media $\mu$ y varianza $\sigma^2$. Entonces, si **n** es suficientemente grande, $\overline{X}$ tiene aproximadamente una distribución normal con

$$\mu_{\overline{x}} = \mu \qquad \text{y} \qquad \sigma^2_{\overline{x}} = \sigma^2/n \;\rightarrow\; \sigma_{\overline{x}} = \sigma / \sqrt{n}$$

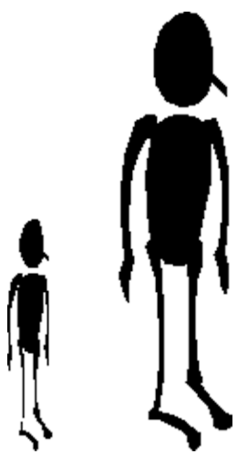
Cuanto más grande sea el tamaño de las muestras **(n)**, mejor será la aproximación.

El Teorema Central del Límite garantiza una distribución Normal cuando "n" es suficientemente grande.

*En resumen, la variable aleatoria en estudio X y la distribución de las medias muestrales siguen distribuciones normales de parámetros:*

| Variable $X$ | Variable medias muestrales $\overline{X}$ |
|---|---|
| $X \rightarrow N(\mu, \sigma)$ | $\overline{X} \rightarrow N(\mu, \sigma / \sqrt{n})$ |

## ACTIVIDAD PROPUESTA 4



Se ha realizado un estudio estadístico sobre la estatura de los 728 alumnos de un IES, comprobándose que se distribuyen normalmente con una media de 171 cm y una desviación típica de 11 cm.

Si encargamos a cada uno de nuestros 33 alumnos de una clase que recojan aleatoriamente muestras de 50 alumnos:

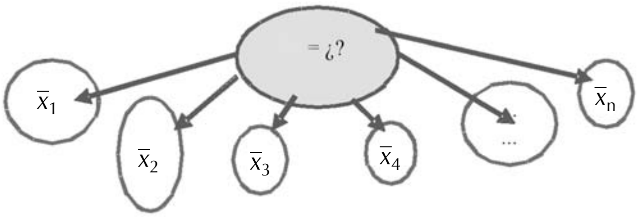
(a) ¿Cuál es la media y la desviación típica esperada de la distribución de las medias muestrales?

(b) ¿En cuántas muestras se espera que la media supere los 173 cm?

(c) ¿En cuántas muestras se espera una media inferior a 170 cm?

(d) ¿En cuántas muestras se espera que la media se encuentre entre 174 cm y 170 cm?

**Propuesta de resolución apartado (a)**



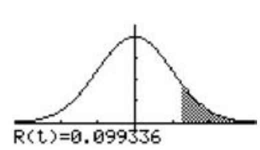

- Variable → $\overline{X}$ ≡ Medias muestrales de las estaturas.
- Ajustes → Como n > 30 aplicamos el Teorema Central del Límite que afirma que $\overline{X}$ se aproxima a una distribución N $\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

$$\mu_{\overline{x}} = \mu = 1.71 \; ; \; \sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} = \frac{11}{\sqrt{50}} = 1.556$$

$\overline{X}$ se aproxima a una distribución N(171; 1.556)

**Propuesta de resolución apartado (b)**

- Se tipifica mediante el cambio $Z = \frac{\overline{X} - 171}{1.556}$
- PLANTEAMIENTO → P[$\overline{X}$ ≥ 173] → $P\left[Z \geq \frac{173-171}{1.556}\right]$

| | |
|---|---|
| Después de realizadas las operaciones aritméticas | 173-171<br>2<br>Ans÷1.556<br>1.285347044<br>MAT |
| SHIFT (Sketch) F4 (GRPH) F5 (Y=) F1 OPTN (▷) F6 (PROB) F3 (▷) F6 (R() F3<br>SHIFT (Ans) (–) | 173-171<br>2<br>Ans÷1.556<br>1.285347044<br>Graph Y=R(Ans<br>P( Q( R( t( |
| EXE |  |

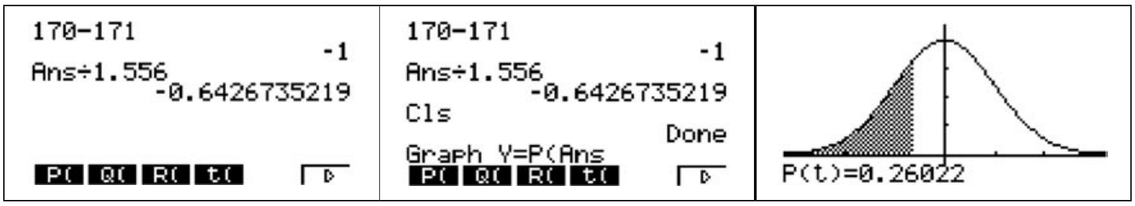
P[$\overline{X}$ ≥ 173] = 0.099336

Como son 33 muestras → 33· 0.099336 = 3.378088

Se espera que la media supere los 173 cm en 3 de las 33 muestras

**Propuesta de resolución apartado (c)**

• PLANTEAMIENTO → $P[\overline{X} \leq 170]$ → tipificamos $P\left[Z \leq \dfrac{170-171}{1.556}\right]$



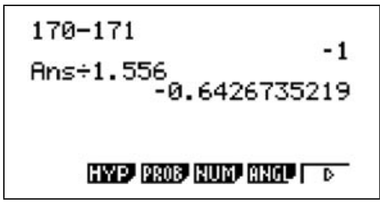

$$P[\overline{X} \leq 170] = 0.26022$$

Como son 33 muestras → 33· 0.26022= 8.58726

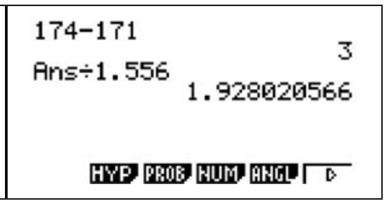
*Se espera que la media sea inferior a 170 cm en 8 de las 33 muestras.*

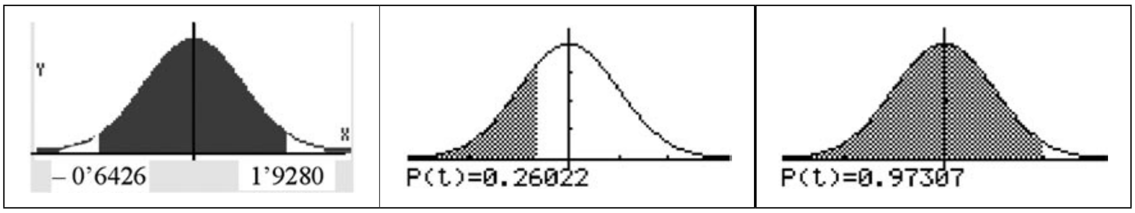
**Propuesta de resolución apartado (d)**

• PLANTEAMIENTO → $P[170 \leq \overline{X} \leq 174]$ → tipificamos $P\left[\dfrac{170-171}{1.556} \leq Z \leq \dfrac{174-171}{1.556}\right]$




Diseñamos la estrategia adecuada para su cálculo




$$P[Z \leq 1.928020566] - P[Z \leq -0.6426735219] =$$

$$= 0.97307 - 0.26022 =$$

$$= 0.71285$$

Como son 33 muestras → 33· 0.71285 = 25.52405

*Se espera que la media se encuentre entre 170 cm y 174 cm en 25 de las 33 muestras.*

## DISTRIBUCIÓN MUESTRAL DE PROPORCIONES

Cuando una variable aleatoria en estudio puede tomar solamente dos valores diferentes (éxito o fracaso) es necesario estimar una proporción o porcentaje. En estos casos o situaciones afirmamos que la población en estudio sigue una distribución Binomial. Cuando el tamaño de la población es grande, n > 30, la distribución Binomial se aproxima a una distribución Normal.


Llamamos p al parámetro poblacional proporción de uno de los valores que presenta la variable en estudio y q al parámetro del otro valor de la variable, siendo q = 1 – p.

Si consideramos todas las muestras de tamaño n que pueden extraerse de la población, de cada muestra obtenemos un valor para la proporción o estadístico muestral de la variable estudiada. Estos valores constituyen una nueva variable aleatoria $\hat{p}$.

Las relaciones que existen entre los parámetros de la población y los estadísticos de la distribución muestral de proporciones son las siguientes:

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

Aplicando el teorema central de límite para muestras grandes tenemos que la distribución Binomial se aproxima a una Normal de parámetros:

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

En la práctica ocurre que las proporciones p y q de la población son desconocidas. En estos casos se aproximan a las obtenidas a partir de una muestra siempre que el tamaño de ésta sea mayor que 100 (n > 100).

## ACTIVIDAD PROPUESTA 5

Un profesor con 120 alumnos tiene 48 aprobados. Si tomamos una muestra de 30 alumnos. ¿cuál es la probabilidad de que haya más de 10 aprobados?

**Propuesta de resolución**

- VARIABLE $\rightarrow \hat{P} \equiv$ Proporciones muestrales.
- AJUSTES $\rightarrow$ n = 30 $\rightarrow$ n ≥ 30 $\rightarrow$ $\hat{P}$ se aproxima a una distribución:

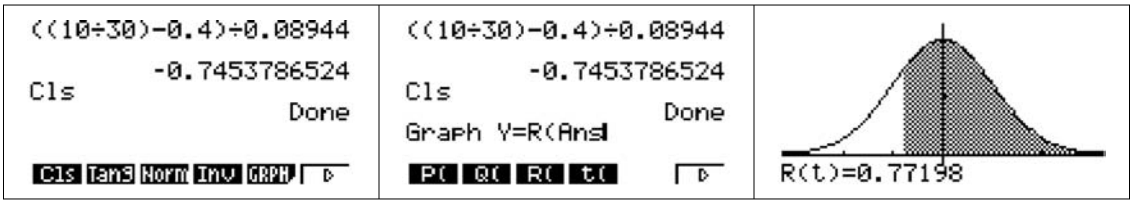
$$N\left(p,\ \sqrt{\frac{p \cdot q}{n}}\right)$$

$$p = \frac{48}{100} = 0.4 \quad \rightarrow \quad 1 - p = 0.6 \quad \rightarrow \quad \sigma_{\bar{x}} = \sigma / \sqrt{n} = 0.08944$$

Las muestras de tamaño 30 extraídas de esa población se ajustan a N(0.4; 0.0894).

Se tipifica mediante el cambio $Z = \dfrac{\hat{p} - 0.4}{0.08944}$

• PLANTEAMIENTO → $P\left[\hat{p} \geq \frac{10}{30}\right]$ tipificamos $P\left[\hat{p} \geq \frac{\frac{10}{30} - 0.4}{0.08944}\right]$




*La probabilidad de que haya más de 10 aprobados, si tomamos una muestra de 30 alumnos, es de 0.77198.*

## PROPUESTAS DE ACTIVIDADES INDAGATORIAS TIPO TEST

[NOTA: *Se puede utilizar como herramienta auxiliar cualquier tipo de calculadora*.]

**1. En el muestreo aleatorio estratificado las muestras de cada estrato son:**

- ☐ Todas iguales.
- ☐ Dependen del tipo de afijación.
- ☐ Proporcionales al numero de elementos de cada estrato.
- ☐ Todas distintas.
- ☐ Ninguna de las afirmaciones anteriores es cierta.

**2. En un muestreo aleatorio estratificado proporcional, la muestra $n_i$ de cada estrato es igual a:**

- ☐ $n_i = \frac{n}{N} \cdot N$,
- ☐ $n_i = \frac{N}{n} \cdot N$,
- ☐ $n_i = \frac{N_i}{N} \cdot N$,
- ☐ $n_i = \frac{N_i}{N} \cdot N$
- ☐ Ninguna de las anteriores.

**3. En una población de tamaño K= 500 dividida en tres estratos de tamaño $N_1 = 150$, $N_2 = 150$ y $N_3 = 250$, se toma una muestra de tamaño n = 15:**

- ☐ Por afijación $n_1 = 6$
- ☐ Por conglomerados $n_3 = 15$
- ☐ Por afijación proporcional $n_2 = 9$

**4. En la población P = {1, 2, 3}, se toman muestras de tamaño 2 sin reemplazamiento. Obtenemos una nueva variable aleatoria media muestral, cuya desviación típica toma el valor:**

- ☐ $\sigma_{\overline{x}} = \frac{1}{\sqrt{3}}$
- ☐ $\sigma_{\overline{x}} = \frac{1}{\sqrt{6}}$
- ☐ $\sigma = \frac{2}{\sqrt{3}}$
- ☐ $\sigma = 2$

**5. En la población P = {2, 3, 4}, estudiamos la variable proporción de cifras impares. Tomamos muestras con reemplazamiento de tamaño 2. La desviación típica de la nueva variable de la proporción muestral es:**

- ☐ $\sigma_p = \frac{1}{\sqrt{3}}$
- ☐ $\sigma_p = \frac{1}{\sqrt{6}}$
- ☐ $\sigma_p = \frac{2}{\sqrt{3}}$
- ☐ $\sigma = 2$

**6. El teorema central del límite nos garantiza que al estudiar la variable $\overline{X}$, (parámetro a estimar):**

- ☐ La distribución de la variable $\overline{X}$ es la misma que la variable en estudio.
- ☐ Está centrada en el mismo lugar que la variable en estudio X.
- ☐ La media de la nueva variable es el doble que la media de la variable en estudio.
- ☐ La nueva variable media muestral sigue una distribución N(0, 1).

**7. Al aplicar el teorema central del límite la nueva variable en estudio $\overline{X}$ sigue una distribución:**

- ☐ $\overline{X} \to N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$
- ☐ $\overline{X} \to N\left(\mu, \frac{\sigma}{n}\right)$
- ☐ $\overline{X} \to N\left(\mu, \frac{\sigma^2}{n}\right)$
- ☐ $\overline{X} \to N\ (1, 0)$

**8. En la distribución muestral de las proporciones, al aplicar el teorema central del límite la nueva variable en estudio ($\hat{P}$) sigue una distribución:**

- ☐ $\overline{X} \rightarrow N\left(p, \dfrac{p \cdot q}{n}\right)$
- ☐ $\overline{X} \rightarrow N\left(p, \sqrt{\dfrac{p \cdot q}{n}}\right)$
- ☐ $\overline{X} \rightarrow N\,(0, 1)$
- ☐ $\overline{X} \rightarrow N\left(\dfrac{p}{n}, \sqrt{\dfrac{p \cdot q}{n}}\right)$