Tema 6: Distribuciones Muestrales

•El objetivo es efectuar una generalización de los resultados de la muestra a la población. Inferir o adivinar el comportamiento de la población a partir del conocimiento de una muestra.

•Para ello es necesario conocer las distribuciones de probabilidad de ciertas funciones de las muestras que constituyen variables aleatorias asociadas al experimento aleatorio, selección de una muestra al azar de una población.

•Estas variables aleatorias denominadas ***estadísticos muestrales***, porque se basan en el comportamiento de las muestras, asignan a cada muestra del espacio muestral, constituido por todas la muestras posibles, un número real que es un ***resumen estadístico de la muestra***. Por ejemplo, media de la muestra.

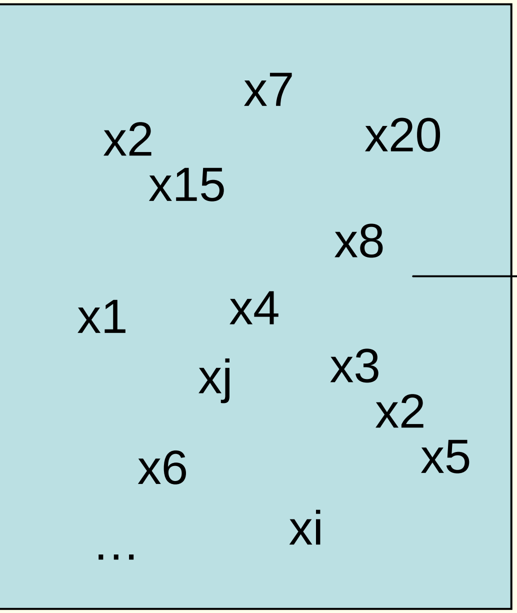
•El esquema siguiente resume visualmente la situación: Dada una población formada por un número N grande de elementos, donde se observa la variable X, se extrae al azar una muestra de tamaño n (n<N).

•Tanto en la población como en la muestra podemos resumir los valores de la variable X observada. Los valores resumen se denominan **parámetros** en la población y **estadísticos** en las muestras. Notaremos con **U** a un estadístico o resumen muestral determinado.

Distribuciones Muestrales

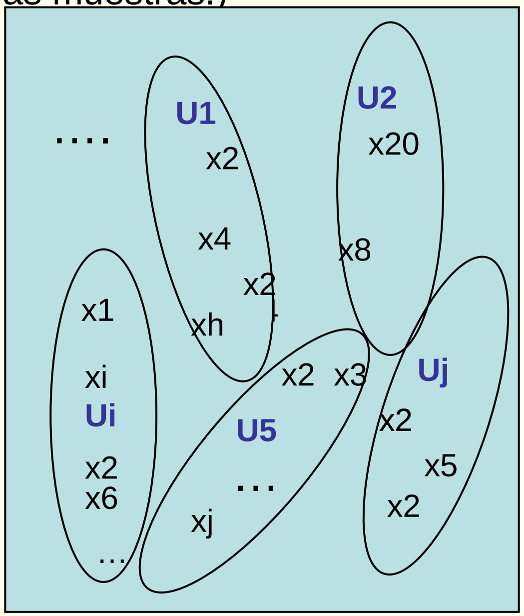
Población formada por todas las muestras posibles de tamaño n
U variable aleatoria muestral
(Estadísticos de las muestras: resúmenes de los valores de X en las muestras.)

Población
X variable aleatoria



Muestra

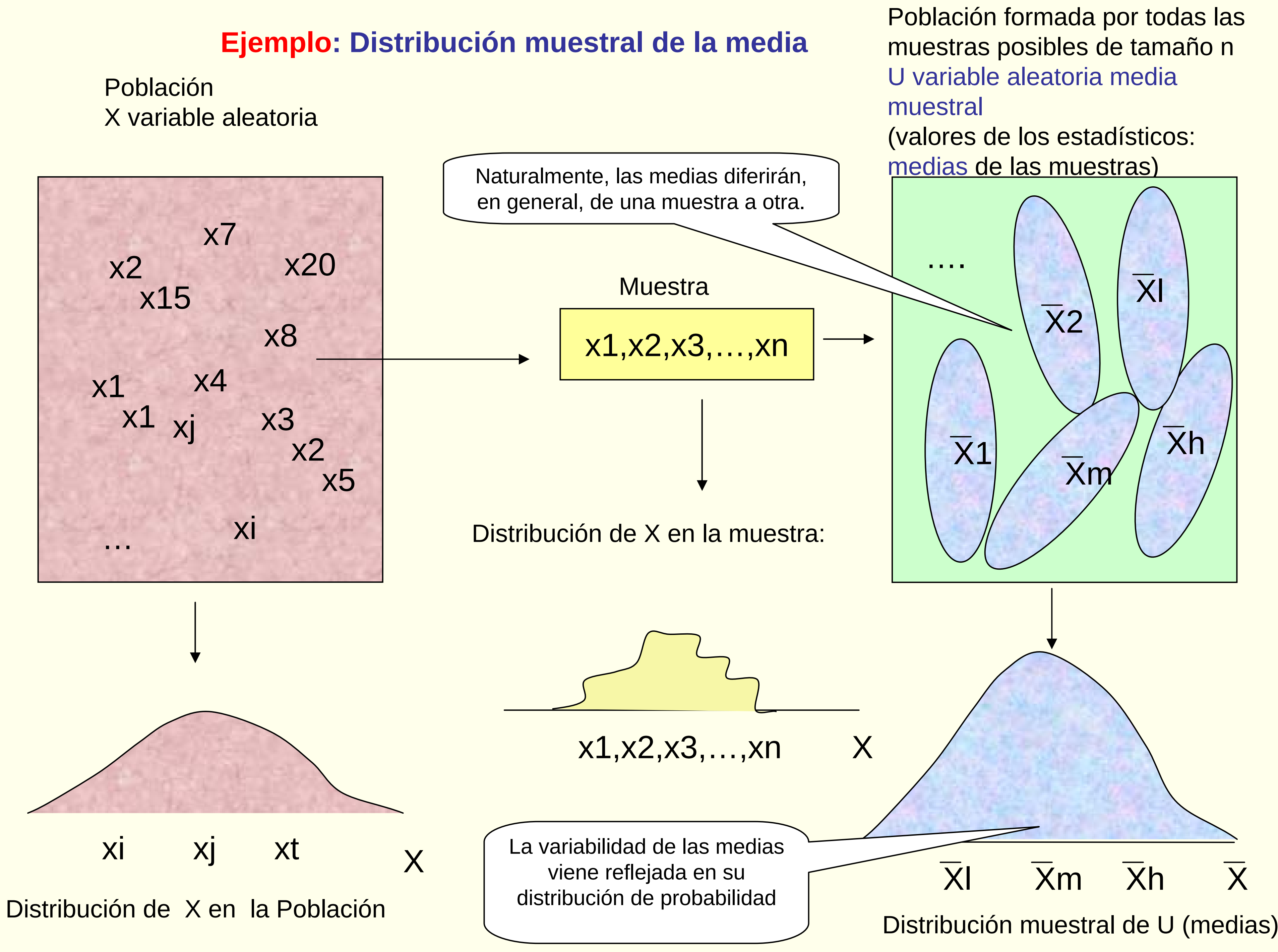
x1,x2,x3,…,xn



Resumen de X en la muestra:
Estadísticos: media, varianza, …

Resumen de X en la Población:
Parámetros: media, varianza, …

Resumen de los estadísticos muestrales:
media, varianza, …


Ejemplo: Distribución muestral de la media
Población
X variable aleatoria
x7
x2
x20
x15
x8
x1
x4
x1
xj
x3
x2
x5
xi
…
Naturalmente, las medias diferirán, en general, de una muestra a otra.
Muestra
x1,x2,x3,…,xn
Distribución de X en la muestra:
Población formada por todas las muestras posibles de tamaño n
U variable aleatoria media muestral
(valores de los estadísticos: medias de las muestras)
….
$\overline{X}$l
$\overline{X}$2
$\overline{X}$1
$\overline{X}$m
$\overline{X}$h
xi
xj
xt
X
Distribución de X en la Población
x1,x2,x3,…,xn
X
La variabilidad de las medias viene reflejada en su distribución de probabilidad
$\overline{X}$l
$\overline{X}$m
$\overline{X}$h
$\overline{X}$
Distribución muestral de U (medias)

Distribuciones Muestrales

- Resumiendo:
- Dada una población y el experimento aleatorio consistente en seleccionar una muestra de dicha población, se define la variable aleatoria U (estadístico muestral) como una aplicación que asigna a cada muestra, **m**, un resumen estadístico determinado,**U(m)**. Esta nueva variable aleatoria U tiene una distribución de probabilidad denominada ***distribución muestral de U***.

$$U : E \longrightarrow R$$

$$m=(x1, x2, \ldots,xn) \longrightarrow U(m)=\text{estadístico muestral}$$

- Su comportamiento dependerá del que tenga X en la población y del tamaño de las muestras.

- Utilidad:

Estaremos interesados en conocer su comportamiento probabilístico, porque esto nos permitirá hacer inferencias acerca del comportamiento de la población.

- A veces nos resultará útil conocer su esperanza matemática y/o su varianza.

Distribuciones Muestrales

•Casos particulares de distribuciones muestrales:

Sea una población P y X la variable aleatoria observada, cuya distribución es Normal

$$X \to N(\mu, \sigma)$$

Total muestral

U=total muestral=t : E $\longrightarrow$ R

m=(x1, x2, …,xn) $\longrightarrow U(m) = t = \sum_{i=1}^{n} Xi$

•Se verifica que el estadístico muestral t es también normal con las siguientes media y desviación típica:

$$\boxed{t = \sum_{i=1}^{n} Xi \to N(n\mu, \sqrt{n\sigma^2})}$$

Además, si el tamaño de la muestra es suficientemente grande (n>30), t se distribuye normalmente, aunque X en la población no sea normal.

Distribuciones Muestrales

•Casos particulares de distribuciones muestrales:

Sea una población P y X la variable aleatoria observada, cuya distribución es Normal

$$X \to N(\mu, \sigma)$$

Media muestral

U=media muestral : E $\longrightarrow$ R

m=(x1, x2, …,xn) $\longrightarrow$ $U(m) = \overline{X} = \dfrac{\sum_{i=1}^{n} Xi}{n}$

•Se verifica que el estadístico muestral media es también normal con las siguientes media y desviación típica:

$$\boxed{\overline{X} = \frac{\sum_{i=1}^{n} Xi}{n} \to N(\mu, \frac{\sigma}{\sqrt{n}})}$$

Además, si el tamaño de la muestra es suficientemente grande (n>30), se distribuye normalmente, aunque X en la población no sea normal.

Distribuciones Muestrales

• Casos particulares de distribuciones muestrales:

Sea una población P y X la variable aleatoria observada, cuya distribución es Normal

$$X \rightarrow N(\mu, \sigma)$$

Estadístico cuasivarianza muestral

$$U = S^2 : E \longrightarrow R$$

$$m = (x1, x2, \ldots, xn) \longrightarrow S^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1}$$

Observa que esta nueva variable se ha obtenido a partir de la anterior (S^2) multiplicando por la constante $(n-1)/\sigma^2$
Esta transformación permite conocer su comportamiento probabilístico

Estadístico Uchi muestral

$$Uchi : E \longrightarrow R$$

$$m = (x1, x2, \ldots, xn) \longrightarrow Uchi(m) = \frac{(n-1)S^2}{\sigma^2}$$

• Se verifica que el estadístico muestral Uchi sigue un modelo Chi-cuadrado con n-1 grados de libertad

$$\boxed{U_{chi} = \frac{(n-1)S^2}{\sigma^2} \rightarrow \chi^2_{n-1}}$$

Distribuciones Muestrales

•Casos particulares de distribuciones muestrales:

Sea una población P y X la variable aleatoria observada, cuya distribución es Normal

$$X \rightarrow N(\mu, \sigma)$$

Estadístico Ut,v (función de los estadísticos media y cuasivarianza)

Ut,v : E ⟶ R

m=(x1, x2, …,xn) ⟶ $U_{t,\upsilon} = \dfrac{\overline{X} - \mu}{\dfrac{S}{\sqrt{n}}}$

Observa que esta nueva variable se ha obtenido a partir de los estadísticos media y cuasidesviación típica (S)
Esta transformación permite conocer su comportamiento probabilístico

•Se verifica que el estadístico muestral Ut,v sigue un modelo t de Student con n-1 grados de libertad

$$\boxed{U_{t,\upsilon} = \frac{(\overline{X} - \mu)}{\dfrac{S}{\sqrt{n}}} \rightarrow t_{n-1}}$$

Distribuciones Muestrales

•Casos particulares de distribuciones muestrales:

Sea una población P y X la variable aleatoria observada, cuya distribución es una Bernoulli:
X=0 con probabilidad q=P(fracaso) y X=1 con probabilidad p=P(éxito)
Se sabe que E(X)=p y V(X)=pq

Estadístico Ut=t=total de éxitos en la muestra

$$Ut : E \longrightarrow R$$

$$m=(x1, x2, \dots, xn) \longrightarrow Ut = t = \sum_{i=1}^{n} Xi$$

Observa que esta variable refleja el total de éxitos (1's) en la muestra entre el total de selecciones n. Es un caso particular de total muestral t, ya visto. Observa que es también la variable binomial

•Se verifica que el estadístico muestral Ut sigue un modelo Binomial

$$\boxed{Ut \rightarrow B(n, p)}$$

E(Ut)=np
V(Ut)=npq

•Para tamaños de muestra suficientemente grandes se aproxima a una normal

Distribuciones Muestrales

•Casos particulares de distribuciones muestrales:

Sea una población P y X la variable aleatoria observada, cuya distribución es una Bernoulli:
X=0 con probabilidad q=P(fracaso) y X=1 con probabilidad p=P(éxito)
Se sabe que E(X)=p y V(X)=pq

Estadístico Up=proporción de éxitos en la muestra

$$Up : E \longrightarrow R$$

$$m=(x1, x2, \ldots, xn) \longrightarrow Up = \frac{Ut}{n} = \frac{\sum_{i=1}^{n} Xi}{n} = \textit{proporción de éxitos}$$

•Si el tamaño de muestra es suficientemente grande (n>30)
se verifica que el estadístico muestral Up sigue un modelo normal

$$\boxed{Up \rightarrow N(p, \sqrt{\frac{pq}{n}})}$$

Observa que esta variable refleja la proporción de éxitos (1's) en la muestra entre el total de selecciones n. Es un caso particular de media muestral , ya visto.

Ejemplo simple de simulación del proceso de generación de una Distribución Muestral de la Media

- Supondremos una población formada por solo N=3 elementos. Por ejemplo 3 niños. Se observa la variable X=edad. Consideremos la selección con reemplazamiento de todas las muestras posibles de tamaño n=2.
- Población niños: {A, B, C} Edad: 2, 3 y 4 años respectivamente.
- El espacio muestral formado por todas las muestras posibles
- E={AA, AB, AC, BA, BB, BC, CA, CB, CC}

- Estadístico muestral U=media muestral

U=media muestral : E $\longrightarrow$ R

m=(x1, x2, …,xn) $\longrightarrow$ $U(m) = \overline{X} = \dfrac{\sum_{i=1}^{n} Xi}{n}$

Ejemplo simple de simulación del proceso de generación de una Distribución Muestral de la Media (continuación)

U=media muestral : E $\longrightarrow$ R

(AA) $\longrightarrow$ $\bar{X} = \frac{2+2}{2} = 2$

(AB) $\longrightarrow$ $\bar{X} = \frac{2+3}{2} = 2{,}5$

(AC) $\longrightarrow$ $\bar{X} = \frac{2+4}{2} = 3$

(BA) $\longrightarrow$ $\bar{X} = \frac{2+3}{2} = 2{,}5$

(BB) $\longrightarrow$ $\bar{X} = \frac{3+3}{2} = 3$

(BC) $\longrightarrow$ $\bar{X} = \frac{3+4}{2} = 3{,}5$

(CA) $\longrightarrow$ $\bar{X} = \frac{2+4}{2} = 3$

(CB) $\longrightarrow$ $\bar{X} = \frac{3+4}{2} = 3{,}5$

(CC) $\longrightarrow$ $\bar{X} = \frac{4+4}{2} = 4$

Ejemplo simple de simulación del proceso de generación de una Distribución Muestral de la Media (continuación)

Distribución muestral de la media

U=media	P(U)
2	1/9
2,5	2/9
3	3/9
3,5	2/9
4	1/9

Distribución de *X=Edad* en la población

X=Edad	P(X)
2	1/3
3	1/3
4	1/3

E(X)=3
Varianza(X)=2/3

Comprueba que la esperanza de X en la población coincide con la esperanza de la variable media muestral, U, y que la varianza de U es igual al cociente entre la Varianza de X y el tamaño de la muestra (2).

Ejemplo: Distribución Muestral de la Media

Observa que en la práctica se selecciona una muestra de la población y se efectúa un resumen de dicha muestra mediante el cálculo de un estadístico muestral que nos interese. Por ejemplo media de la muestra, desviación típica, mediana, etc.

Si conociéramos el comportamiento que tienen todos los posibles valores del estadístico muestral que nos interesa (su modelo de probabilidad), podríamos saber qué probabilidad hay de que el valor de nuestra muestra esté comprendida en un determinado intervalo.

Ejemplo:

Se ha seleccionado una muestra al azar de 50 mujeres de una población de mayores de 18 años. Se desconoce la talla media de la población, pero en la muestra se ha observado que la media de las 50 tallas es 1,60 m. Si se sabe por otros estudios que la desviación típica en la población es de 3 ,3 cm, determina la probabilidad de que la media de la población no difiere en más de 1 cm de la de la muestra.

Ejemplo: Distribución Muestral de la Media

Ejemplo (continuación):

Se sabe que para tamaños de muestra grandes la media muestral se distribuye según un modelo normal.

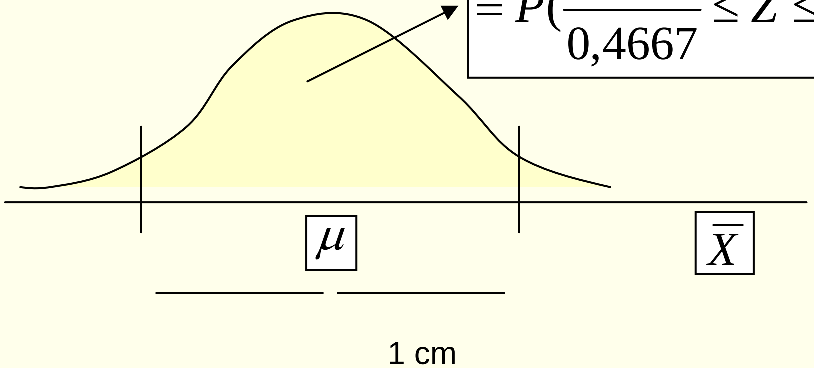
$$\overline{X} = \frac{\sum_{i=1}^{n} Xi}{n} \rightarrow N(\mu, \frac{\sigma}{\sqrt{n}})$$

Sustituyendo

$$\overline{X} \rightarrow N(\mu, \frac{3{,}3}{\sqrt{50}}) \equiv N(\mu, 0{,}4667)$$

$$P(|\overline{X} - \mu|) < 1) = P(\mu - 1 \leq \overline{X} \leq \mu + 1)$$

$$= P(\frac{-1}{0{,}4667} \leq Z \leq \frac{1}{0{,}4667}) = P(-2{,}14 \leq Z \leq 2{,}14)$$



Ejemplo: Distribución Muestral de la Media

Ejemplo 2

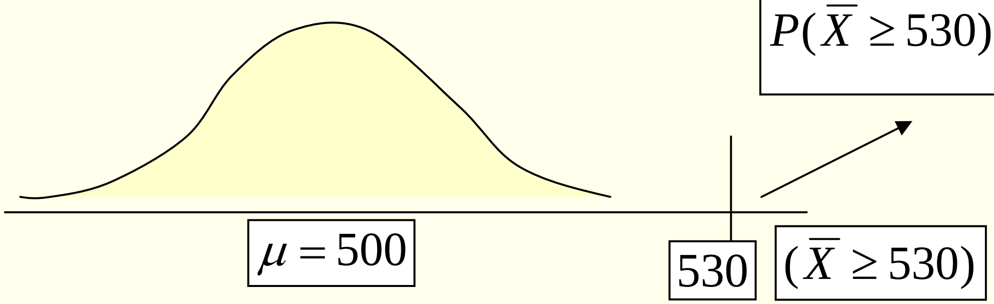
Se sabe que los pesos de los paquetes de cierto artículo en una cadena de producción se distribuyen normalmente con media 500 gr y desviación típica 10 gr. Se selecciona una muestra de 100 paquetes de la producción y se observa que la media de éstos es de 530 gr ¿es coherente este resultado con la hipótesis de que se distribuyen normalmente con media y desviación típica 500 y 10, respectivamente.

Si la hipótesis $\boxed{X \rightarrow N(\mu = 500, \sigma = 10)}$ fuese cierta, entonces habrá que admitir

$$\boxed{\overline{X} \rightarrow N(500, \frac{10}{\sqrt{100}}) \equiv N(500, 1)}$$

La probabilidad de observar un suceso tan extremo o más (530 gr) es igual a

$$\boxed{P(\overline{X} \geq 530) = P(Z \geq \frac{530 - 500}{1}) = P(Z \geq 30) = 0}$$



Ejemplo: Distribución Muestral de la Media

Ejemplo 3

Se sabe que los pesos de los paquetes de cierto artículo en una cadena de producción se distribuyen normalmente con media 500 gr y desviación típica 10 gr
Se seleccionan muestras de 10, 50 y 100 paquetes de la producción.
Determina la desviación típica de la media muestral para cada uno de estos 3 casos.

$$X \to N(\mu = 500, \sigma = 10)$$

Caso n=10

$$\bar{X} \to N(500, \frac{10}{\sqrt{10}}) \equiv N(500, 3{,}16)$$

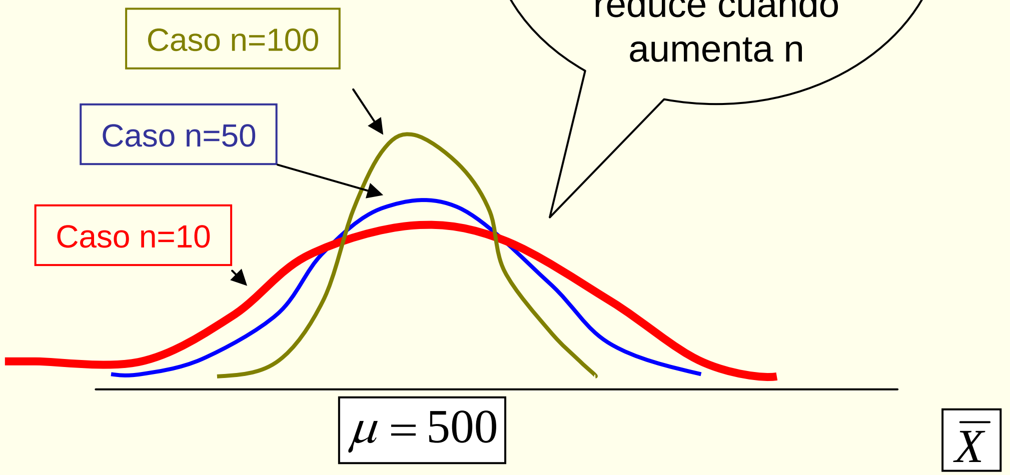
Caso n=50

$$\bar{X} \to N(500, \frac{10}{\sqrt{50}}) \equiv N(500, 1{,}41)$$

Caso n=100

$$\bar{X} \to N(500, \frac{10}{\sqrt{100}}) \equiv N(500, 1)$$

Observa que la variabilidad de la distribución muestral se reduce cuando aumenta n



Ejemplo: Distribución Muestral de la Media

Ejemplo 4

En unas elecciones un determinado candidato asegura que tiene ganados al menos el 50% de los votos. En un sondeo previo a las elecciones se obtuvo una muestra de 500 votantes y 160 se mostraron favorables al candidato ¿es coherente con la hipótesis del candidato el resultado obtenido en la muestra?

Si la hipótesis $\boxed{X \to B(1, p = 0{,}5)}$ fuese cierta, entonces habría que admitir que la proporción muestral se distribuye según

$$\boxed{Up \to N\left(p, \sqrt{\frac{pq}{n}}\right) \equiv N\left(0{,}5, \sqrt{\frac{0{,}5 \cdot 0{,}5}{500}}\right) \equiv N(0{,}5, \quad 0{,}1118)}$$

La probabilidad de observar un suceso tan extremo o más al obtenido Up=160/500=0,32 es

$$\boxed{P(U_p \le 0{,}32) = P\left(Z \le \frac{0{,}32 - 0{,}5}{0{,}1118}\right) = P(Z \le -16{,}1) = 0}$$