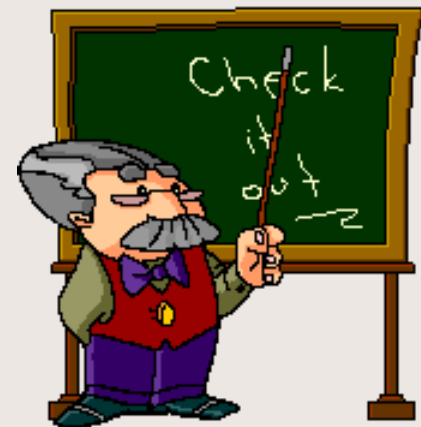


# Distribuciones Muestrales



## Teorema Central del Límite

## Distribuciones Muestrales

Para hacer inferencias sobre una población, es necesario examinar un poco más los resultados muestrales.

Suponer que se toma una muestra aleatoria de  $n$  elementos y se calcula la media aritmética ( $\bar{X}_n$ ) ¿Coincidirá con  $\mu$  ?

Si se toma una segunda muestra ¿tendrá la media igual a la de la muestra anterior? ¿Coincidirá con  $\mu$  ?

Es de esperar que los valores estén “próximos” al valor poblacional.

Pero ¿Qué se entiende por “próximos” ? ¿Cómo se determina o mide ésta proximidad o cercanía?

¿Cómo se distribuye un estadístico muestral cuando se ha muestreado repetidamente?

Un estadístico muestral es una variable aleatoria.

# Medidas Estadísticas Descriptivas

Una distribución muestral es aquella distribución teórica de probabilidades de los valores posibles de un estadístico muestral que ocurriría si obtuviéramos todas las muestras posibles de un tamaño fijo, de una variable correspondiente a una población dada.

## Estadísticos muestrales importantes:

Sea  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$  una muestra aleatoria de una variable cuantitativa  $\mathbf{X}$ , entonces las variables:

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \qquad T_n = \sum_{i=1}^n X_i$$

$$S_X^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

Si  $Y$  es una variable aleatoria con distribución binomial

$$Y \sim \text{Bin}(n, p)$$

$$\frac{Y}{n} = \text{“la proporción muestral de éxitos”}.$$

## Distribuciones Muestrales

Una distribución muestral es aquella distribución teórica de probabilidades de los valores posibles de un estadístico muestral que ocurriría si obtuviéramos todas las muestras posibles de un tamaño fijo, de una variable correspondiente a una población dada.

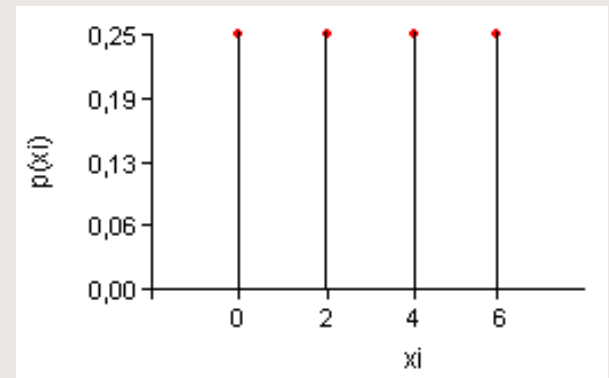
Ejemplo:  $X=\{0; 2; 4; 6\}$  ← es la Distribución poblacional de  $X$

$x_i$	$p(x_i)$	$x_i p(x_i)$	$x_i^2 p(x_i)$
0	0,25	0	0
2	0,25	0,5	1
4	0,25	1	4
6	0,25	1,5	9
tot		3	14

$$E(X) = 3$$

$$V(X) = E(X^2) - [E(X)]^2$$

$$V(X) = 14 - 9 = 5$$



Construyan todas las muestras posibles de tamaño 2  
Calcular esperanza y varianza.

## Distribuciones Muestrales

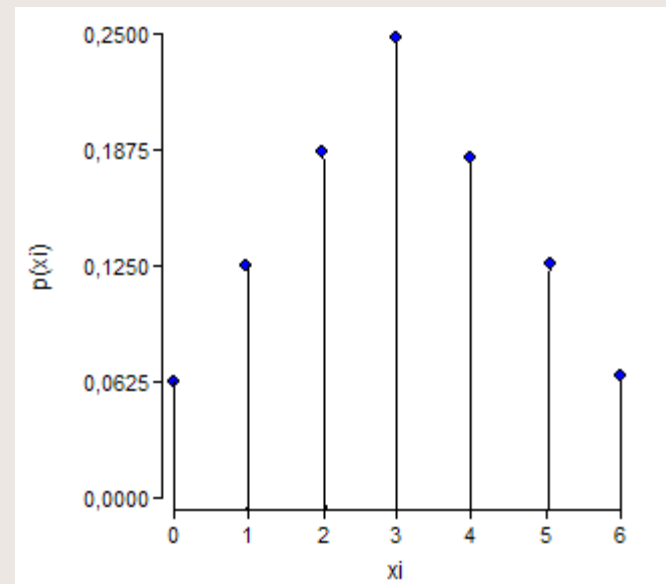
Ejemplo:  $X=\{0; 2; 4; 6\}$  ← es la Distribución poblacional de  $X$

$y_i$	$p(y_i)$	$y_i p(y_i)$	$y_i^2 p(y_i)$
0	0,063	0	0
1	0,125	0,125	0,125
2	0,188	0,375	0,75
3	0,25	0,75	2,25
4	0,188	0,75	3
5	0,125	0,625	3,125
6	0,063	0,375	2,25
tot		3	11,5

Si  $n=2$  y llamamos  $Y$  a la media  
 $E(Y)= 3$

$$V(Y)= E(Y^2) - [E(Y)]^2$$

$$V(Y)= 11,5 - 9 = 2,5 = 5/2$$



**Ayuda para encontrar todos los pares posibles:** colocamos los números en la primer fila y la primer columna

	0	2	4	6
0	(0,0)	(0,2)	(0,4)	(0,6)
2	(2,0)	(2,2)	(2,4)	(2,6)
4	(4,0)	(4,2)	(4,4)	(4,6)
6	(6,0)	(6,2)	(6,4)	(6,6)

Tabla con el promedio de cada par:

	0	2	4	6
0	0	1	2	3
2	1	2	3	4
4	2	3	4	5
6	3	4	5	6

# Distribuciones Muestrales

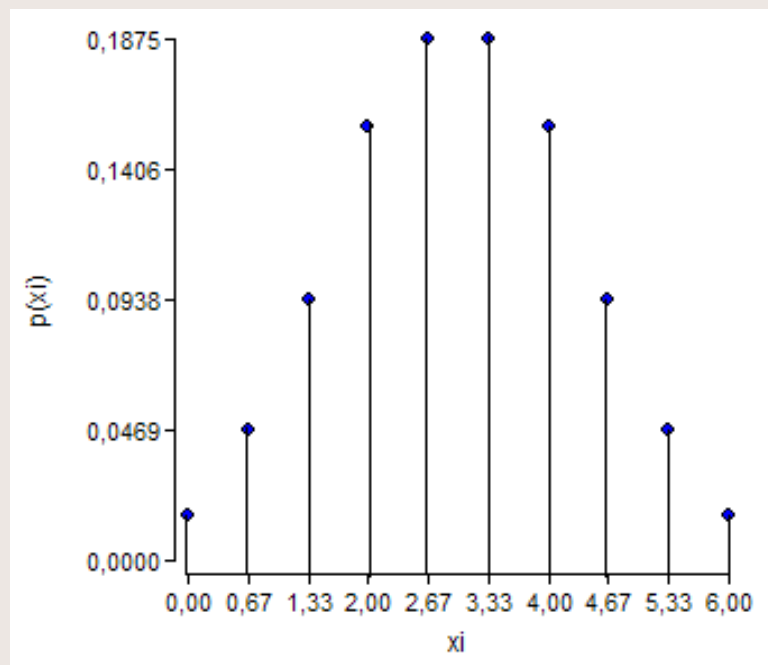
Ejemplo:  $X=\{0; 2; 4; 6\}$  ← es la Distribución poblacional de  $X$

$y_i$	$p(y_i)$	$y_i p(y_i)$	$y_i^2 p(y_i)$
0	0,016	0	0
0,666667	0,047	0,03125	0,020833
1,333333	0,094	0,125	0,166667
2	0,15625	0,3125	0,625
2,666667	0,1875	0,5	1,333333
3,333333	0,1875	0,625	2,083333
4	0,15625	0,625	2,5
4,666667	0,09375	0,4375	2,041667
5,333333	0,047	0,25	1,333333
6	0,016	0,09375	0,5625
tot		3	10,66667

Si  $n=3$  y llamamos  $Y$  a la media  
 $E(Y)=3$

$$V(Y) = E(Y^2) - [E(Y)]^2$$

$$V(Y) = 10,66 - 9 = 1,66 = 5/3$$





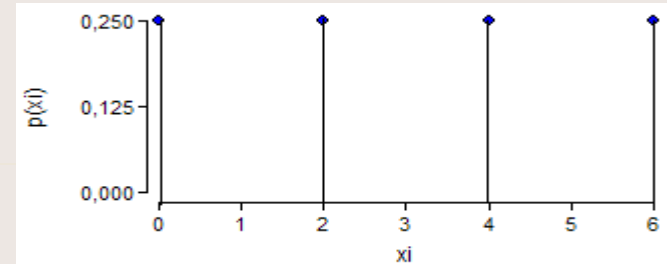
# Distribuciones Muestrales

Cuando  $n=1$

$$E(\bar{X}_1) = 3$$

$$V(\bar{X}_1) = E(\bar{X}_1^2) - [E(\bar{X}_1)]^2 = 14 - 9 = 5/1$$

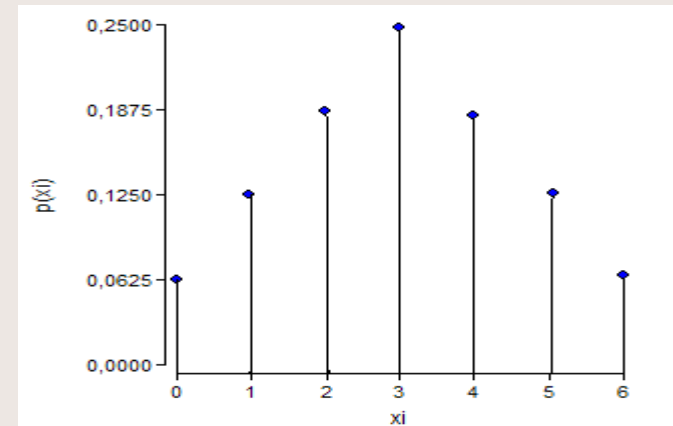
$$X = \{0; 2; 4; 6\}$$



Cuando  $n=2$

$$E(\bar{X}_2) = 3$$

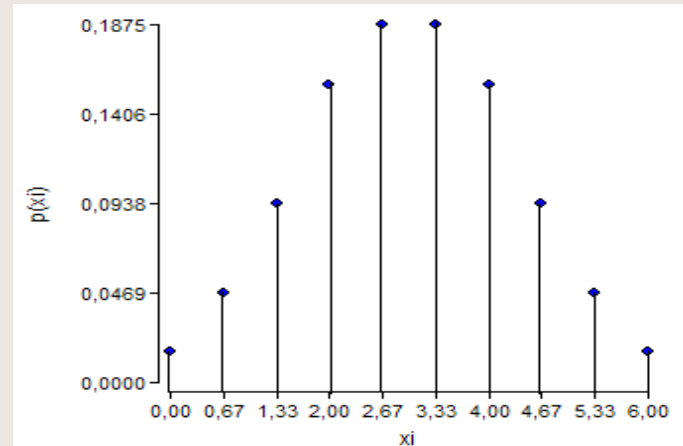
$$V(\bar{X}_2) = E(\bar{X}_2^2) - [E(\bar{X}_2)]^2 = 11,5 - 9 = 5/2$$



Cuando  $n=3$

$$E(\bar{X}_3) = 3$$

$$V(\bar{X}_3) = E(\bar{X}_3^2) - [E(\bar{X}_3)]^2 = 10,66 - 9 = 5/3$$



## Distribución de la media Muestral

Conclusión:

- La media de la distribución muestral no\_ varía al cambiar el tamaño de la muestra. (Siempre es  $\mu$ ;  $E( X ) = \mu$  para todo  $n$  )
- La desviación estándar de la media muestral (llamada *error estándar* de la media) indica que la dispersión de la media muestral decrece a media que crece el tamaño de muestra

$$\sigma_{\bar{X}}^2 = \frac{\sigma_X^2}{n}$$

# Estadísticos Muestrales

## Distribución muestral de

$$\bar{X}_n$$

Sea  $X$  una v.a. tal que

$$E(X) = \mu \quad \text{y} \quad V(X) = \sigma^2$$

y una muestra aleatoria de tamaño  $n$

$$X_1, X_2, X_3, \dots, X_n$$

Entonces:

$$\text{a) } E(\bar{X}) = \mu$$

$$\text{b) } V(\bar{X}) = \frac{\sigma^2}{n}$$

$$\text{c) Si } X \sim N(\mu; \sigma^2) \rightarrow \bar{X} \sim N\left(\mu; \frac{\sigma^2}{n}\right)$$

Nota: 1) no importa cual es el tamaño de la muestra; si  $X$  se distribuye *normalmente* entonces *la media aritmética se distribuye normalmente*.

2) Importante: la varianza de la media muestral es siempre menor que la varianza original.

## Teorema Central del Límite

El *Teorema Central del Límite*, es la herramienta más importante de la teoría de distribuciones pues *garantiza las propiedades distribucionales de la media muestral*, la que tiene un rol destacado en la estadística inferencial (estimación y prueba de hipótesis).

Dada una población cualquiera con media  $\mu$  y varianza  $\sigma^2$  la distribución muestral de  $\bar{x}$ , calculada a partir de las muestras aleatorias de tamaño  $n$  con reemplazo de esta población, estará distribuida en forma aproximadamente normal con  $\mu_{\bar{x}} = \mu$  y varianza  $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$  y la aproximación a la normal será mejor cuanto mayor sea  $n$ .

## **¿Qué significa $n$ suficientemente grande?**

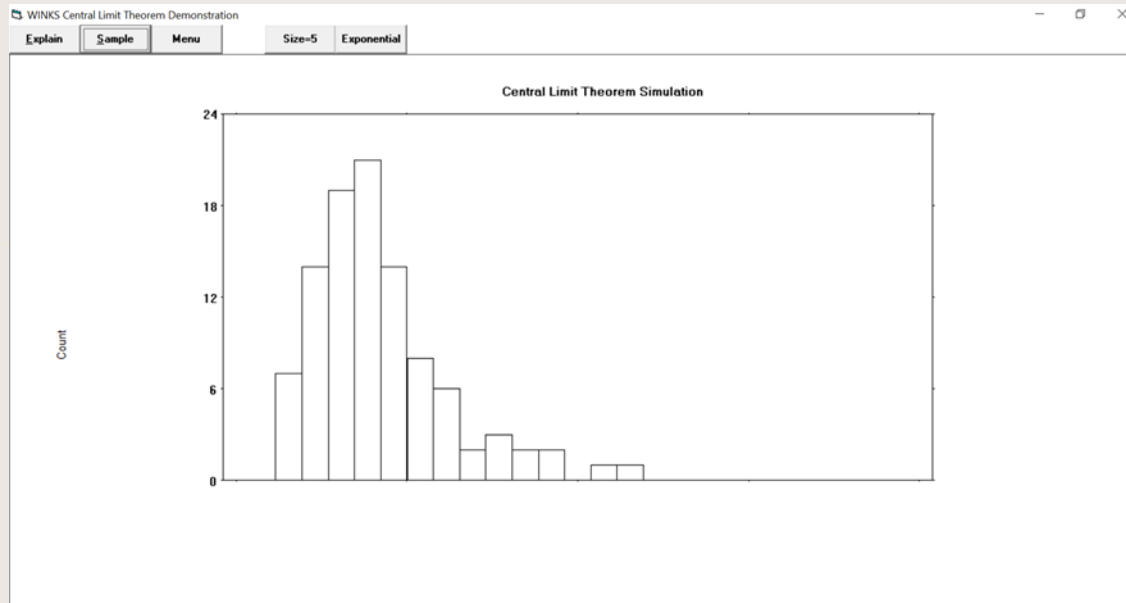
### **¿Cómo sabemos si la aproximación es buena?**

El tamaño  $n$  de la muestra necesario para que la aproximación sea razonable depende la distribución de la variable original. Mientras más acampanada y simétrica es la distribución de las observaciones, más rápidamente se logra una buena aproximación; pero en términos generales basta con  $n \geq 30$  para obtener buenas aproximaciones.

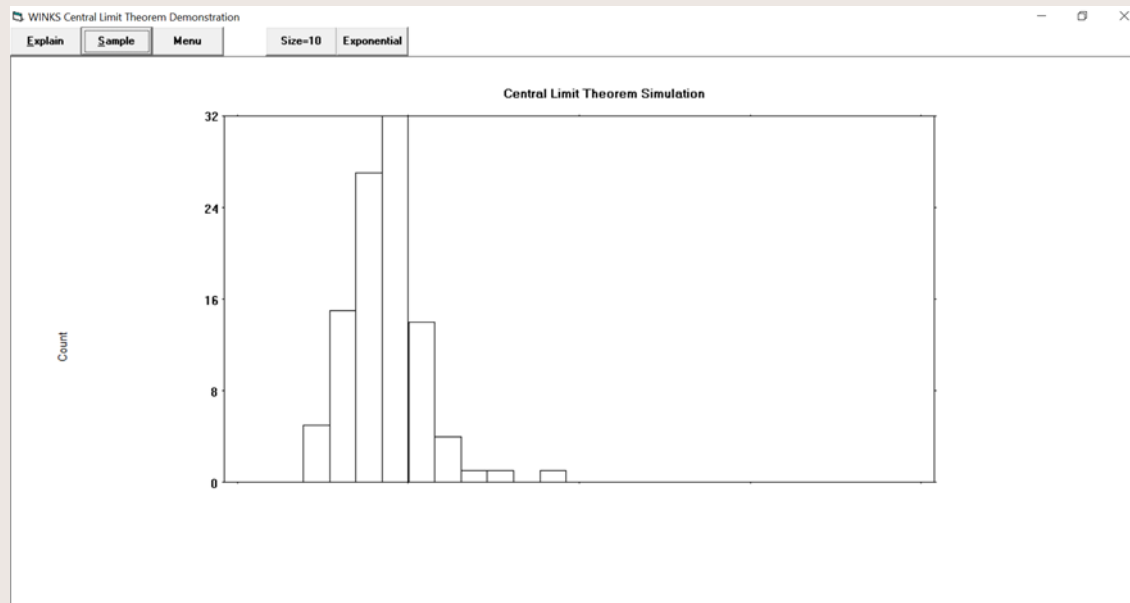
Si de *cualquier población* independiente de la forma de la distribución se extraen muestras aleatorias de tamaño  $n$  fijo, a medida que  *$n$  crece* la distribución de las medias muestrales tiende a convertirse en una distribución *normal* con media  $\mu$  y varianza  $\sigma^2/n$

## Ejemplo- Distribución Exponencial

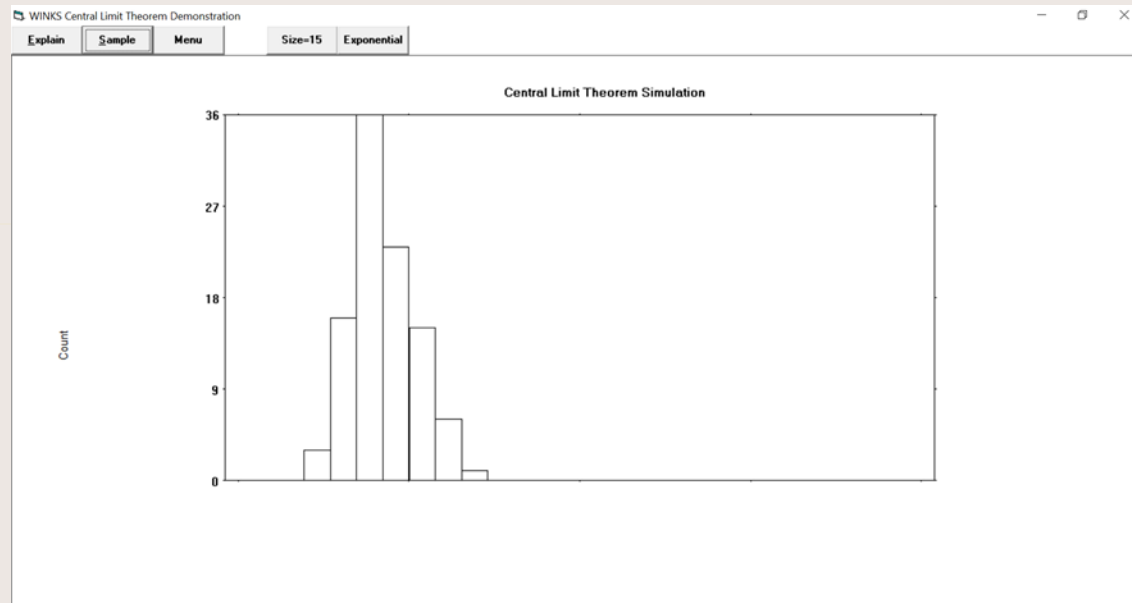
100 medias de muestra de tamaño 5 a 30  
Simulación software Winks SDA



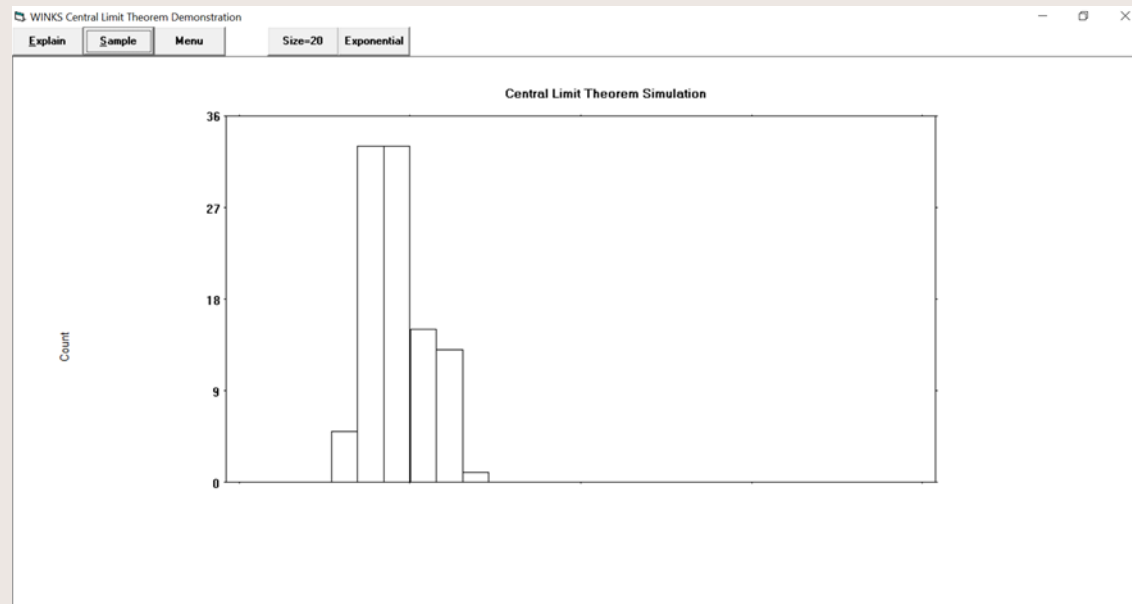
$n=5$



$n=10$

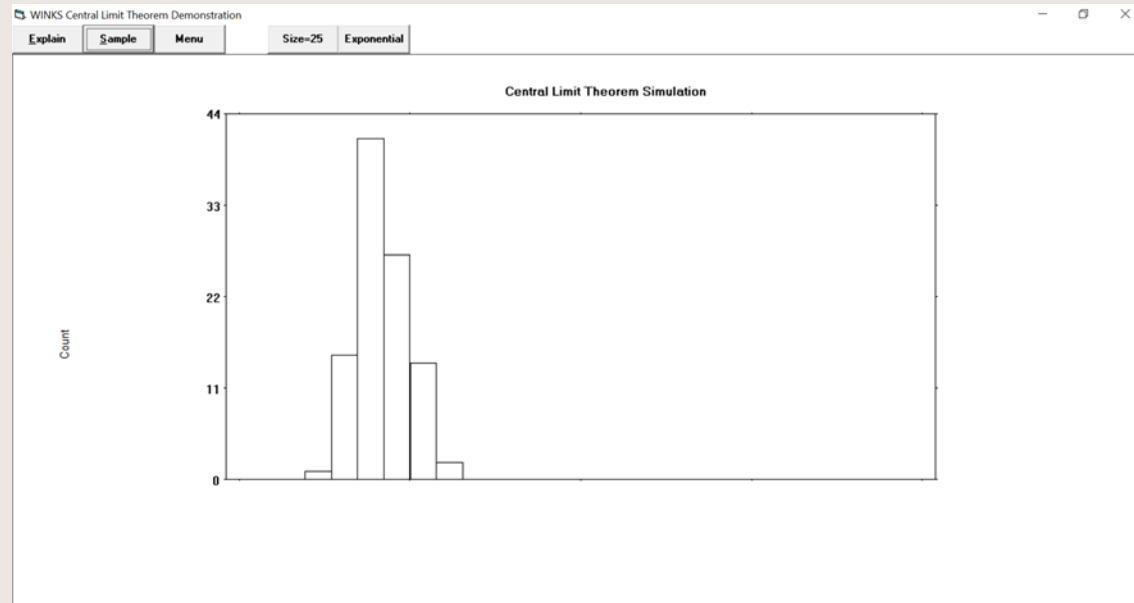


$n=15$

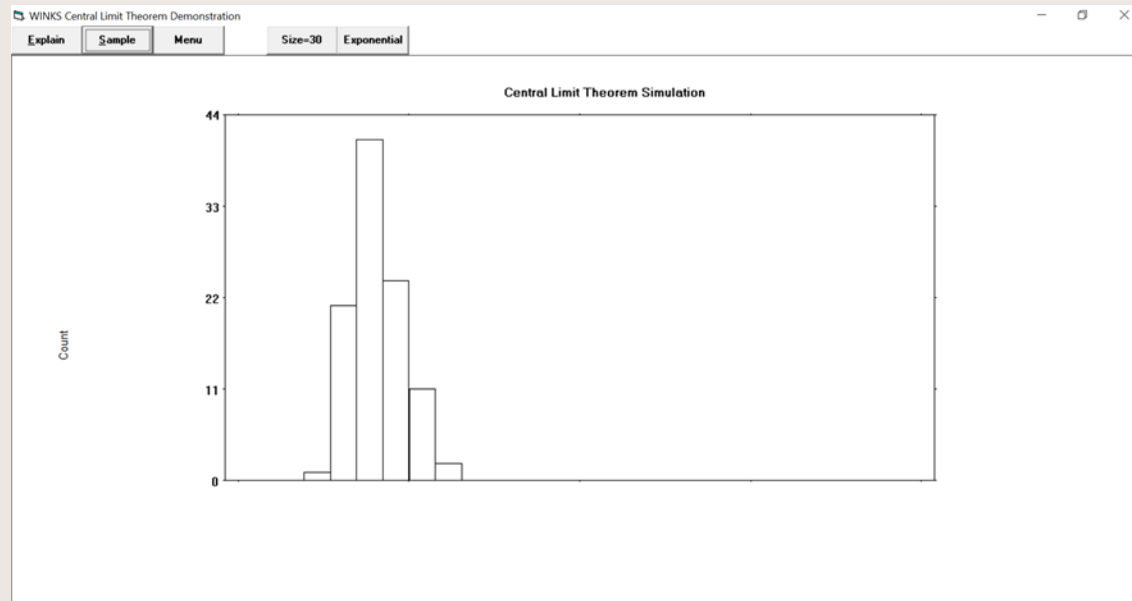


$n=20$





$n=25$



$n=30$

## En el caso de muestras sin reposición

Volvemos al ejemplo original;  $X=\{0; 2; 4; 6\}$

Pares	$\bar{x}$		
0,2	1		
0,4	2		
0,6	3		
2,4	3		
2,6	4		
4,6	5		
tot			

$$E(X) = 3$$

$$V(X_{\text{med}}) = E(X^2) - [E(X)]^2$$

$$V(X_{\text{med}}) = 64/6 - 9 = 5/3$$

Escriba aquí la ecuación.

No hace falta hacer estos cálculos. Si tomamos el valor de la varianza para muestras de tamaño 2 con reemplazo y multiplicarlo por el factor de corrección por población finita y llegamos al valor que calculamos arriba.

$$\sigma_{\bar{X}}^2 = \frac{\sigma_X^2}{n} \left( \frac{N-n}{N-1} \right) = \frac{5}{2} \left( \frac{4-2}{4-1} \right) = \frac{5}{3}$$

# Estadísticos Muestrales

## Distribución muestral de

$$\bar{X}_n$$

Sea  $X$  una v.a. tal que

$$E(X) = \mu \quad \text{y} \quad V(X) = \sigma^2$$

y una muestra aleatoria de tamaño  $n$

$$X_1, X_2, X_3, \dots, X_n$$

Entonces:

a)  $E(\bar{X}) = \mu$

b)  $V(\bar{X}) = \frac{\sigma^2}{n}$

c) Si  $X \sim N(\mu; \sigma^2) \rightarrow \bar{X} \sim N\left(\mu; \frac{\sigma^2}{n}\right)$

Nota: 1) no importa cual es el tamaño de la muestra; si  $X$  se distribuye *normalmente* entonces *la media aritmética se distribuye normalmente*.

2) Importante: la varianza de la media muestral es siempre menor que la varianza original.

### *Ejemplo 1:*

Asuma que el aumento de peso en un cobayo, alimentado con un determinado balanceado, se distribuye normal con media igual a 105 g y desvío estándar igual a 10,5 g. Si se extrae de esta población de cobayos una muestra aleatoria de tamaño 16:

- a) ¿Con qué probabilidad un cobayo tomado al azar tendrá un aumento de peso de al menos 104 g?
- b) ¿Cuál es la probabilidad de que los cobayos de una muestra de tamaño 16 tengan un aumento de peso promedio inferior a 106 g?

a) En este caso estamos calculando la probabilidad de que un cobayo pese 104 g o más (al menos 104 g), nos interesa la distribución de . Entonces se procede:

$$X \sim N(105; 110,25)$$

$$P(X \geq 104) = P\left(Z \geq \frac{104 - 105}{10,5}\right) = P(Z \geq -0,095) = 1 - P(Z < -0,095)$$

buscamos en la tabla de la distribución normal estándar aproximando 0,095 como tenemos sólo 2 decimales tomamos como valor aproximado 0,1 y obtenemos:

$$P(X \geq 104) \cong 1 - P(Z < -0,10) \cong 1 - 0,4602 \cong 0,5398$$

Un cobayo de la muestra tendrá un aumento de peso de al menos 104 g con una probabilidad de 0,54.

**b)** En este caso nos preguntan sobre el aumento medio (o promedio) de los cobayos de una muestra de tamaño 16, entonces se debe usar la distribución de  $\bar{X}_{16}$ .

$$P(\bar{X}_{16} < 106) \equiv P\left(Z < \frac{106 - 105}{\left(\frac{10,5}{\sqrt{16}}\right)}\right) = P(Z < +0,38) = 0,6480$$

La probabilidad de que el promedio de los individuos de una muestra de tamaño 16 sea menor que 106 g es igual a 0,65.

### *Ejemplo 2:*

Una máquina dosificadora está programada para que la cantidad de alimento balanceado que arroja en cada bolsa sea una variable aleatoria con una media de 2000g y una desviación estándar de 150g. ¿Cuál es la probabilidad de que la cantidad media de alimento servido en una muestra aleatoria de 36 bolsas sea por lo menos 2040g ? Realizar los cálculos usando el TCL.

Se tiene  $\mu=2000$     $\sigma=150$     $n=36$

Usando el TCL    $\bar{X} \xrightarrow{d} N\left(\mu, \frac{\sigma^2}{n}\right)$    entonces

$$P(\bar{X} > 2040) = P\left(Z > \frac{2040 - 2000}{150/\sqrt{36}}\right) = P(Z > 1,6) = 0,0548$$

## Corolario del Teorema Central del Límite de De-Moivre- Laplace

Este corolario permite aproximar la distribución binomial por la normal:

Sea  $X \sim \text{Bin}(n;p)$  y sea  $\frac{X}{n}$  = “la proporción muestral de éxitos”.

	Variable	Media	Desvío
<b>Binomial</b>	x	$E(x) = \mu_x = np$	$\sigma_x = \sqrt{npq}$
<b>Proporción</b>	x/n	$E(x/n) = \frac{np}{n} = \pi$	$\text{Desv}(x/n) = \sqrt{\frac{pq}{n}}$

por el Teorema Central del límite:

$$X \approx N(np; np(1-p)) \Rightarrow y \Rightarrow \frac{X}{n} \approx N\left(p; \frac{p(1-p)}{n}\right)$$

$$z \approx \frac{p - \pi}{\sqrt{\frac{p(1-p)}{n}}}$$

Se considera que la aproximación es buena cuando:  $np \geq 5$  y además  $np(1-p) \geq 5$



### ***Ejemplo 3:***

Se ha determinado que 60% de los estudiantes de una universidad grande trabaja más de 6 hs diarias.

Si se toma una muestra aleatoria de 800 de dichos estudiantes, ¿cuál es la probabilidad de que la proporción de los que trabaja más de 6hs en la muestra sea menor que 0,55?

*Solución:*

*Aplicando el TCL sabemos que*

$$p \xrightarrow{d} N \left( p, \frac{p(1-p)}{n} \right)$$

$$Z \approx \frac{p - \pi}{\sqrt{\frac{p(1-p)}{n}}}$$

$$P(p < 0,55) = P \left( Z < \frac{0,55 - 0,60}{\sqrt{\frac{0,60 \cdot 0,40}{800}}} \right) = P(Z < -2,89) = 0,00193$$

## Recordemos

Parámetro $\theta$	Estimador $\hat{\theta}$
$\mu$	$\bar{X}$
$\sigma^2$	$S^2$
$\pi$	$p$

Un estimador es un valor que puede calcularse a partir de los datos muestrales y que proporciona información sobre el valor del parámetro. Un estimador es una variable aleatoria.

$\hat{\theta}$  estimador de  $\theta$

# PROPIEDADES DE LOS ESTIMADORES

1.-Se dice que un estimador es **insesgado** si  $E[\hat{\theta}] = \theta$

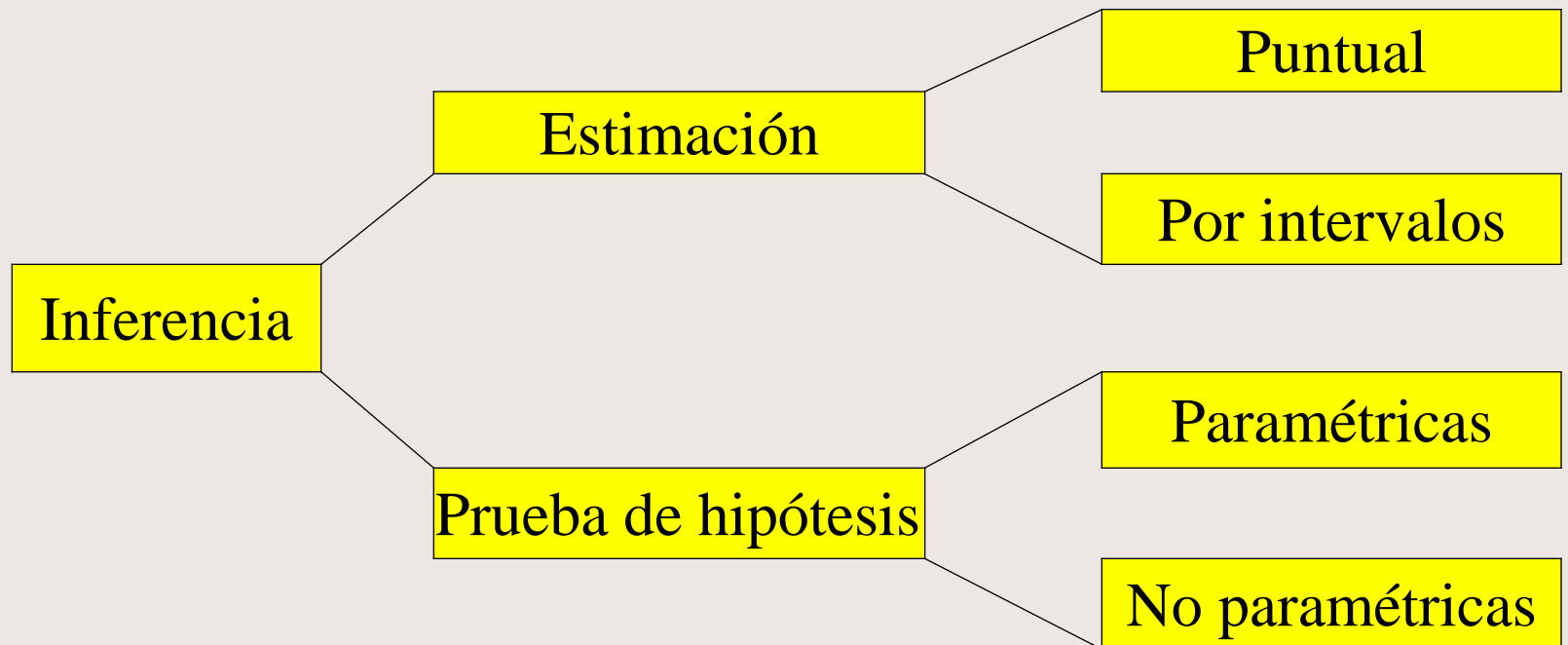
2.-El estimador  $\hat{\theta}_1$  es más **eficiente** que  $\hat{\theta}_2$  si

$$Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$$

3.-Un estimador es **consistente** si  $\lim_{n \rightarrow \infty} \hat{\theta} = \theta$

4.-Un estimador suficiente para un parámetro es aquel capaz de recoger o resumir toda la información que la muestra de una Variable aleatoria  $X$  contiene.

# INFERENCIA



**En la práctica, interesa no sólo dar una estimación de un parámetro, sino que además, un intervalo que permita precisar la incertidumbre existente en la estimación.**

# Gracias

