Unidad III

Muestras Aleatorias
Distribución de Muestreo

Parte II

Estadístico:

Media muestral

$\bar{X}$????

Proposición

Sea X ~ **N(μ, σ^2),** y **X_1, X_2, ..., X_n** una muestra aleatoria de la v.a. X, ie, ***n*** variables aleatorias **independientes normalmente** distribuidas, con $E(X_i) = \mu$ y $V(X_i) = \sigma^2$ finita, i = 1,2, ..., *n*. Entonces la distribución de la v.a. **media muestral**, $\overline{X}$ es **Normal** con media **μ** y varianza **σ^2/n**.

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

En resumen:

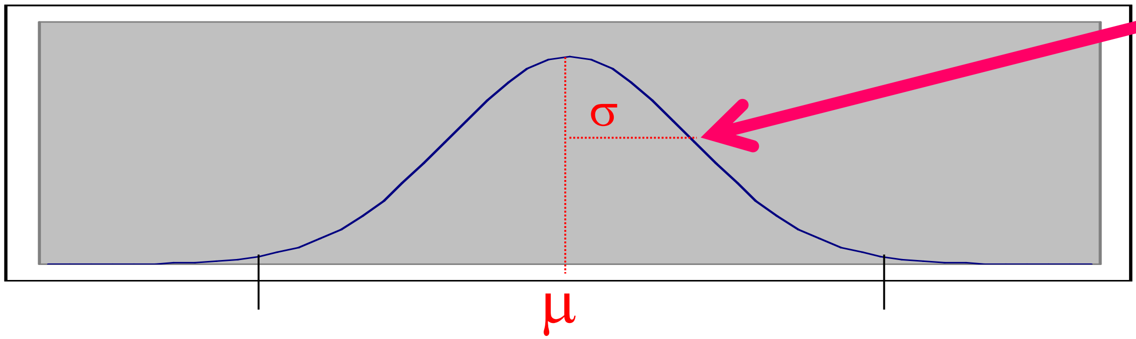
Si X ~ N(μ, σ^2) entonces $\overline{X}$ ~ N(μ, σ^2 /n).

Toda combinación lineal de v.a. independientes Normales es un v.a. con distribución Normal

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{1}{n} X_1 + \ldots + \frac{1}{n} X_n$$

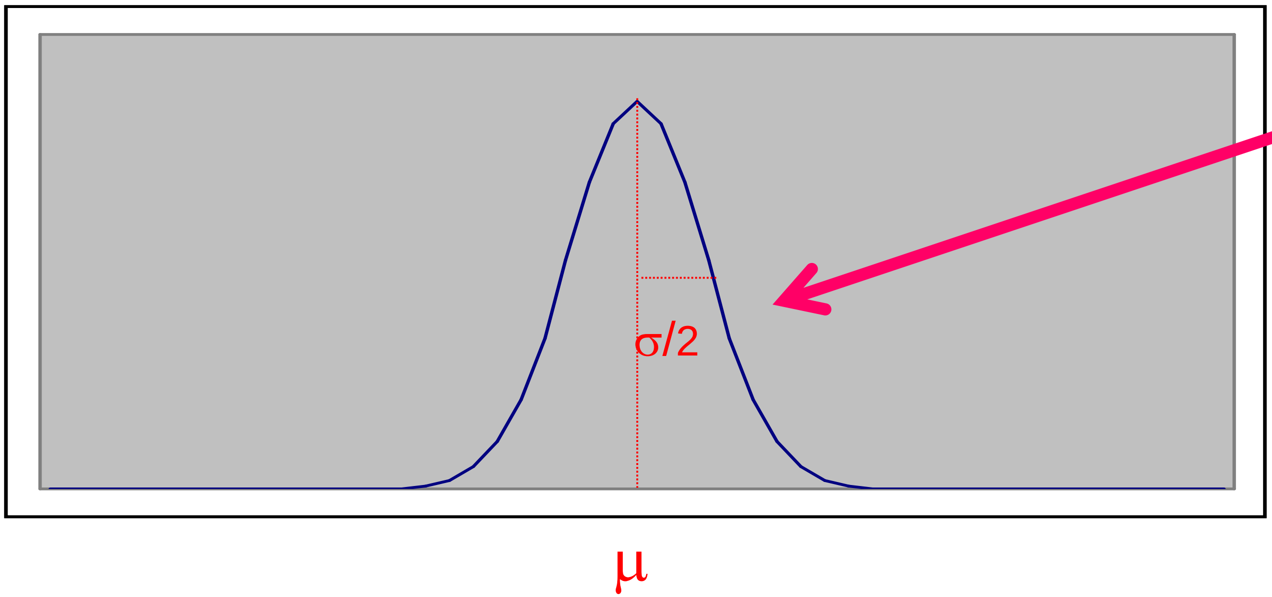
Distribución de la población, X

$X \sim N(\mu, \sigma^2)$



Distribución de la media muestral, n = 4

$\overline{X} \sim N(\mu, \sigma^2 /4)$



Ejemplo

La Cámara de Comercio de Buenos Aires ha registrado de años anteriores que la cantidad promedio de dinero que gasta la gente que asiste a convenciones en comidas, alojamiento y entretenimiento por día es de **5670** pesos, con un desvío de **750** pesos. Suponiendo que la cantidad de dinero gastada en un día es una **v.a.** distribuida normalmente, si de las distintas convenciones que se llevan a cabo en la ciudad, se seleccionaron **16** personas al azar, ¿cuál es la probabilidad de que el gasto promedio sea:

a) menor que 5220 pesos?

b) entre 5400 y 5760 pesos?

c) Supere los 6000 pesos?

Distribución de muestreo

a) ¿cuál es la probabilidad de que el gasto promedio sea menor que 5220 pesos?

Sea **X** = "*cantidad de dinero gastada en un día en comidas, alojamiento y entretenimiento por persona* " . La v.a., **X** $\sim N(\mu = 5670, \sigma^2 = 750^2)$.

$\overline{X}$ = "*cantidad de dinero promedio gastado en un día en comidas, alojamiento y entretenimiento por 16 personas seleccionadas al azar*".

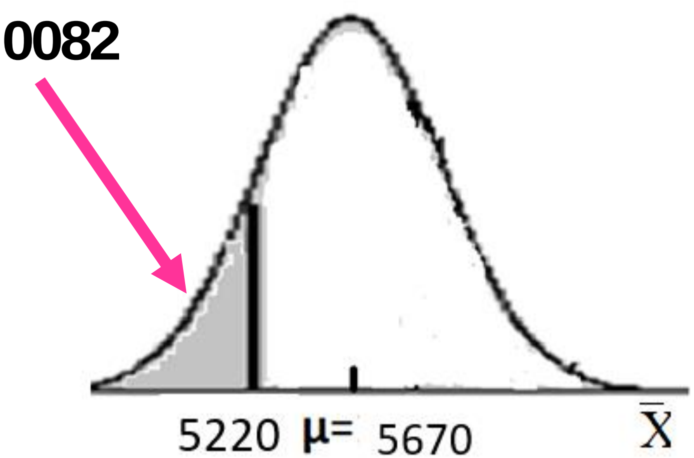
Por la proposición anterior,

como **X** $\sim$ N ($\mu = 5670$, $\sigma^2 = 750^2$)   )

$\overline{X} \sim N(\mu = 5670, \sigma^2 / n = 750^2/16)$

$$P(\overline{X} < 5220) = \mathbf{0.0082}$$

0.0082



En el **0.82%** de las muestras de tamaño 16 el dinero promedio gastado por día en comidas, alojamiento y entretenimiento es inferior a 5220$.

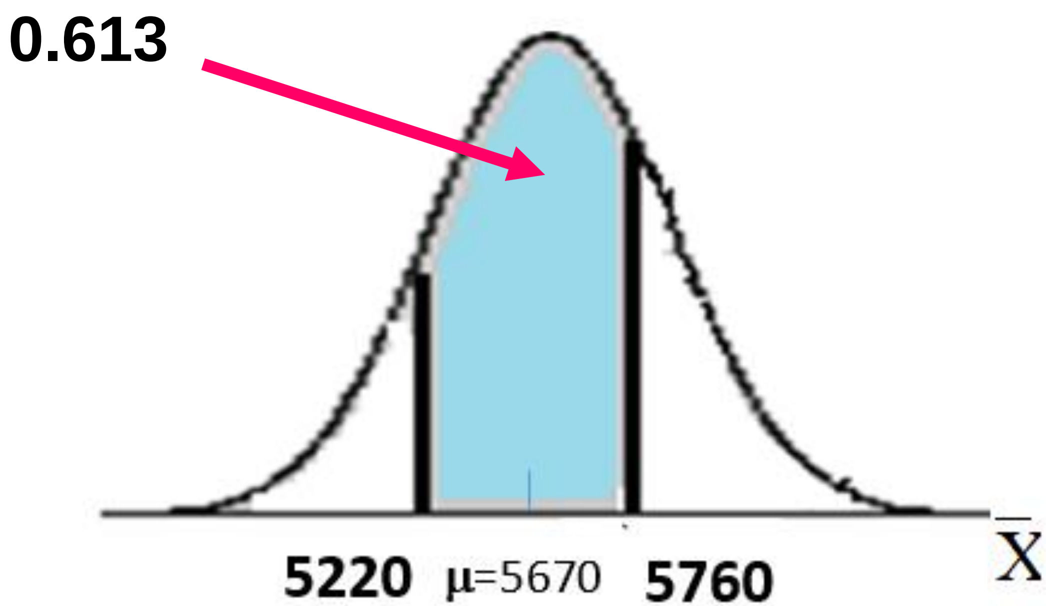
Si de las distintas convenciones que se llevan a cabo en la ciudad, se seleccionaron 16 personas al azar, **¿cuál es la probabilidad de que el gasto promedio sea:**

b) entre 5400 y 5760 pesos? $\bar{X} \sim N(\mu = 5670, \sigma^2 / n = 750^2/16)$

$$P(5400 < \bar{X} < 5760) =$$

$$= P(X < 5760) - P(X < 5400)$$

$$= 0.6879 - 0.0749 = \mathbf{0.613}$$

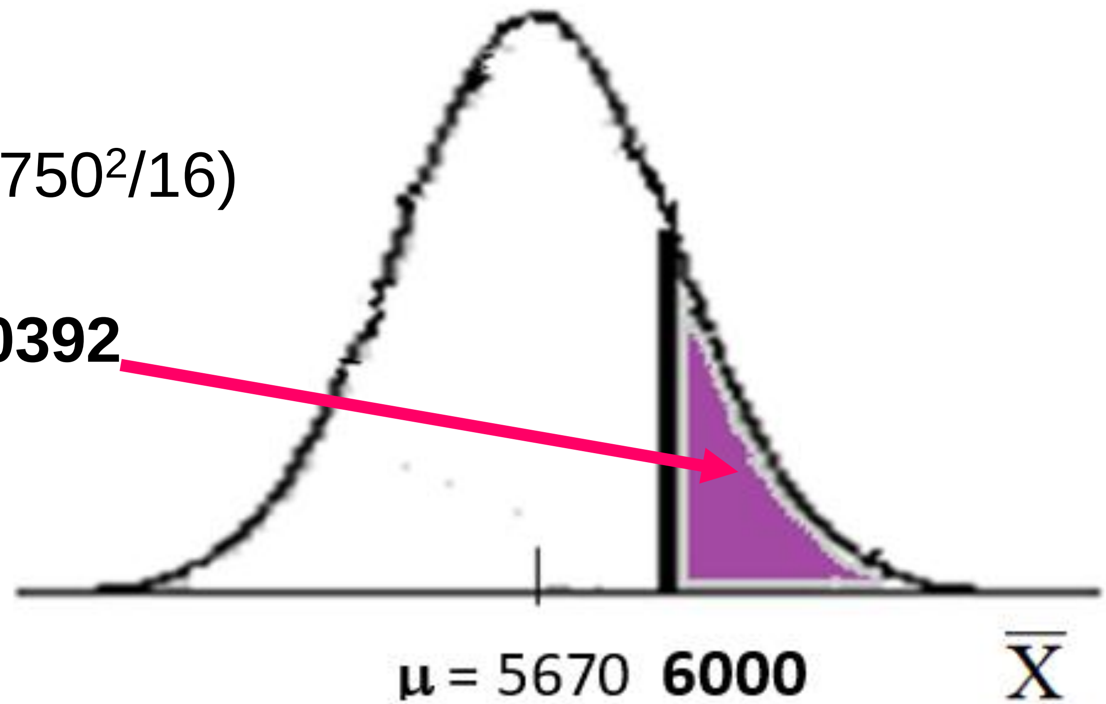


En el **61.3%** de las muestras de tamaño 16 el dinero promedio gastado por día en comidas, alojamiento y entretenimiento oscila entre 5400$ y 5760$.

Si de las distintas convenciones que se llevan a cabo en la ciudad, se seleccionaron 16 personas al azar, **¿cuál es la probabilidad de que el gasto promedio sea:**

c) Supere los 6000 pesos?

$\overline{X} \sim N\ (\mu = 5670,\ \sigma^2 / n = 750^2/16)$



$$P(\overline{X} > 6000) =$$

$$= \mathbf{0.0392}$$

En el **3.92%** de las muestras de tamaño 16 el dinero promedio gastado por día en comidas, alojamiento y entretenimiento supere los 6000$.

Ejercicios

1. La temperatura, X, en cierta región sigue una distribución normal con media 20 ºC y desvío 4 ºC. Si se eligen al azar n días.

a) Determinar la distribución con sus parámetros de la temperatura media muestral, si

i) $n = 4$ **ii)** $n = 20$ **iii)** $n = 500$

b) Para $n = 4$, ¿cuál es la probabilidad de que la temperatura media muestral, no exceda los 18.6 ºC?

2. En un servicio de atención al cliente, el tiempo de espera hasta recibir atención es una v.a. normal de media 10 minutos y desvío estándar de 2 minutos. Se toman muestras aleatorias del tiempo de espera de los clientes que llegan un día concreto. Se pide:

a) ¿Cuál es la probabilidad de que el tiempo medio de espera de una muestra de 25 clientes no supere los 9 minutos?

b) ¿Cuál es la distribución de la media muestral, si se toman muestras aleatorias de 64 clientes? Especificar sus parámetros.

3. La duración de las baterías de un determinado modelo de teléfono móvil tiene una distribución normal de media 34,5 horas y desviación estándar 6,9 horas. Se toma una muestra aleatoria simple de 36 teléfonos móviles.

a) ¿Cuál es la probabilidad de que la duración media de las baterías de la muestra esté comprendida entre 32 y 33,5 horas.
b) ¿Y de que sea mayor de 38 horas?

4. Una fábrica de autos lanza al mercado el modelo "Mathe" del que se sabe que su peso medio es de 3100 kilos con una desviación de 130 kilos.

a) ¿Qué distribución seguirá el peso medio de las muestras aleatorias de 100 autos Mathe?
b) ¿Cuál será la probabilidad de que al comprar 100 coches el peso promedio de ellos sea de más de 2900 kilos y menos de 3500?

Distribución de muestreo

Si se muestrea una población **que tiene una distribución de probabilidad desconocida** la distribución de la media muestral seguirá siendo aproximadamente normal, con media μ y varianza σ^2/n, si el tamaño de la muestra, ***n***, **es grande**. Este es uno de los teoremas más útiles en estadística, se le conoce como **Teorema Central del Límite**, y se enuncia de la siguiente manera:

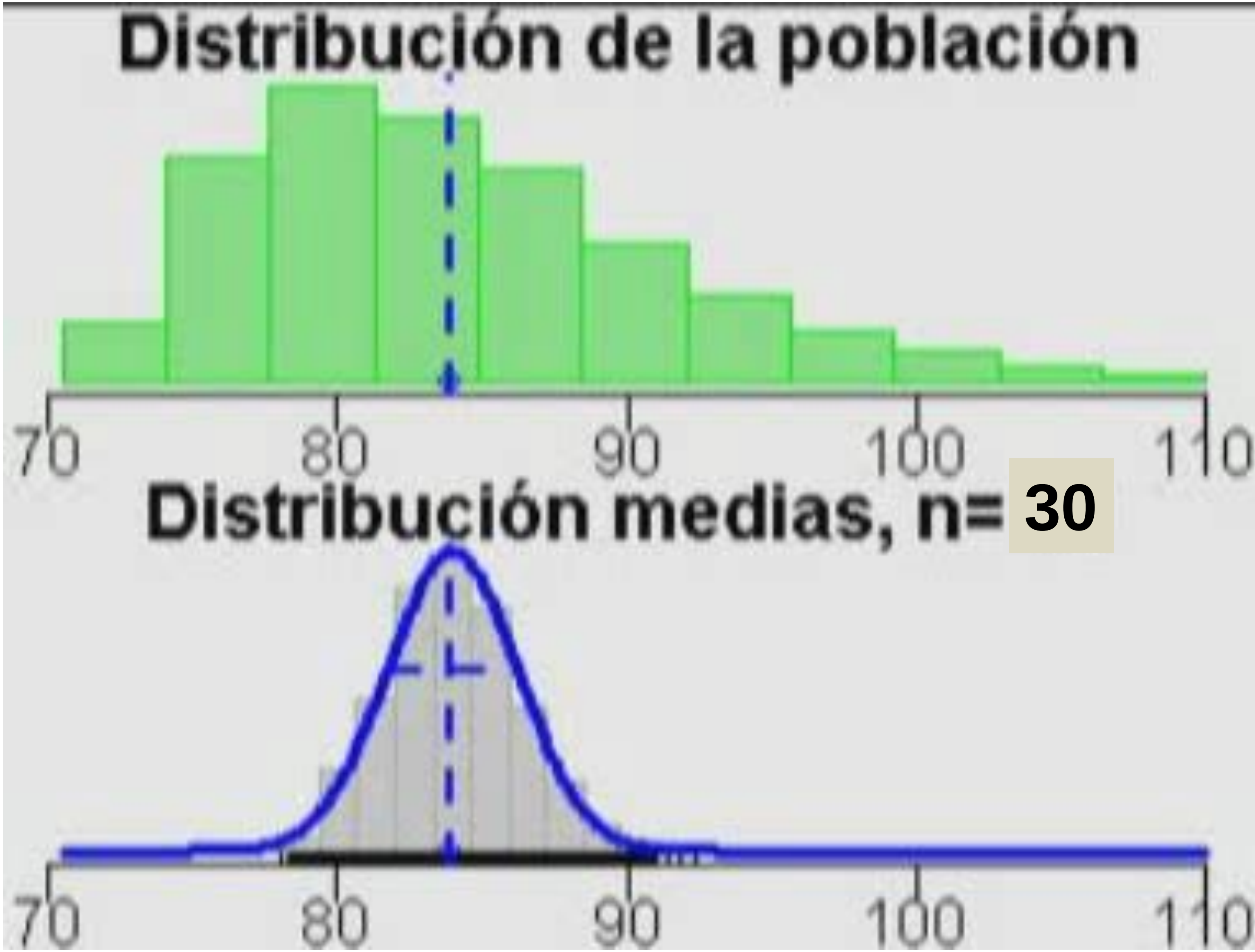
Teorema Central del Límite

Si $X_1, X_2, \ldots, X_n$ es una muestra aleatoria de tamaño *n* de una v.a. X, con media μ y varianza finita σ^2, y si $\overline{X}$ es la media muestral, entonces

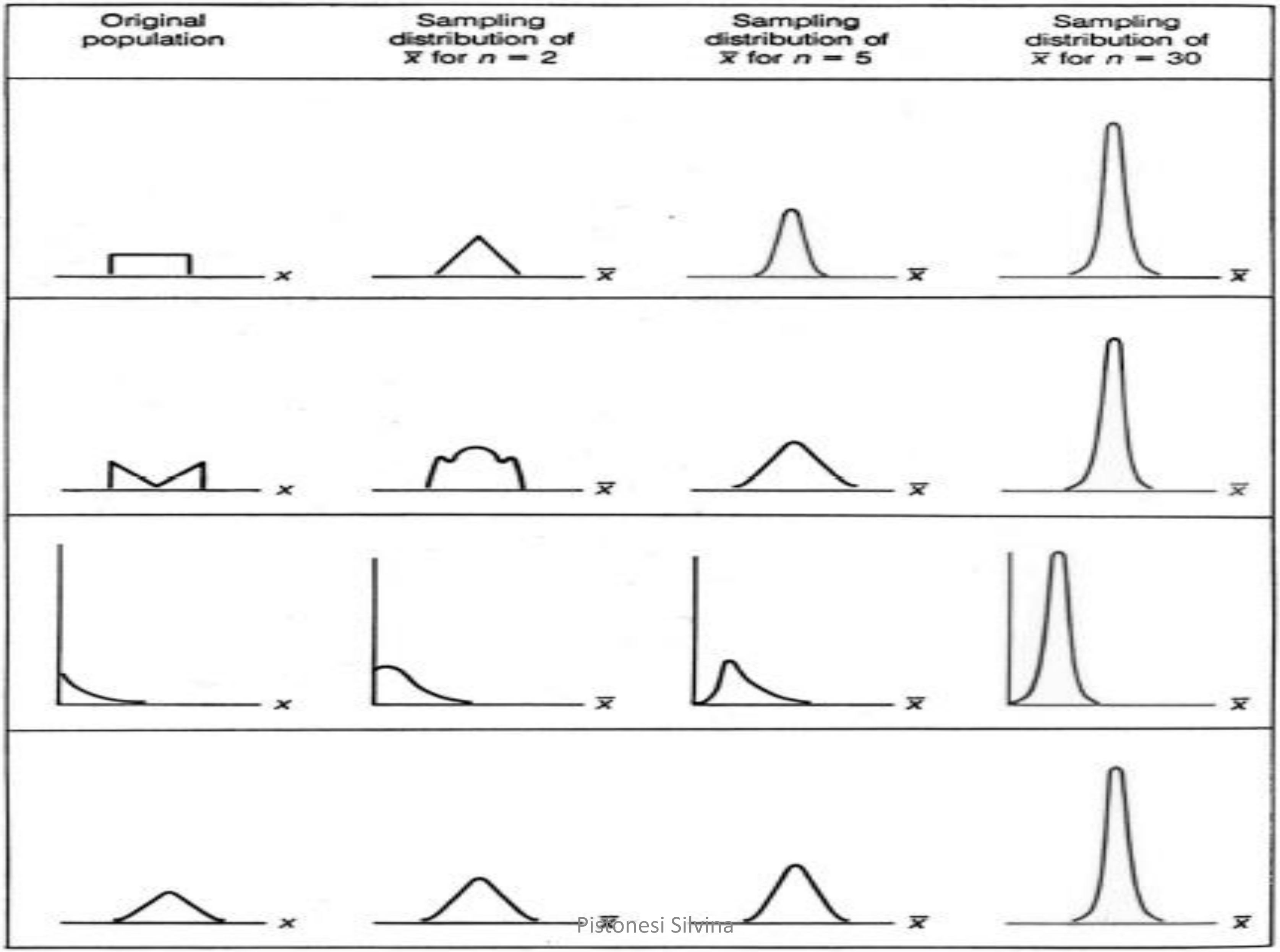
cuando n $\rightarrow \infty$,

$$\overline{X} \sim N(\mu, \sigma^2/n) \quad \text{y} \quad \Rightarrow \quad Z = \frac{\overline{X} - \mu}{\sqrt{\sigma^2/n}} \sim N(0,1).$$

La esencia de **Teorema Central del Límite** recae en que para tamaños de muestra, ***n*, grandes**, la distribución de $\overline{X}$ es, **aproximadamente normal** con media μ y varianza σ^2 **/n,** sin importar cual sea el modelo de probabilidad a partir del cual se obtuvo la muestra.


Distribución de la población
70
80
90
100
110
Distribución medias, n= 30
70
80
90
100
110

Distribución de muestreo



Ejemplo 1

Una empresa piensa comprar una gran cantidad de procesadores a cierto fabricante. Éste asegura a la empresa que la eficacia promedio de este tipo de procesadores para un cierto tipo de tareas, medida a través del tiempo de CPU, es de 48.2 seg. con un desvío estándar igual a 15 seg.

La empresa decide comprar los procesadores sólo si en una muestra aleatoria de **36** de éstos, el tiempo de CPU **promedio** es de a los sumo 42.3 seg.

¿Cuál es la probabilidad de que la empresa adquiera los procesadores?

Ejemplo 1

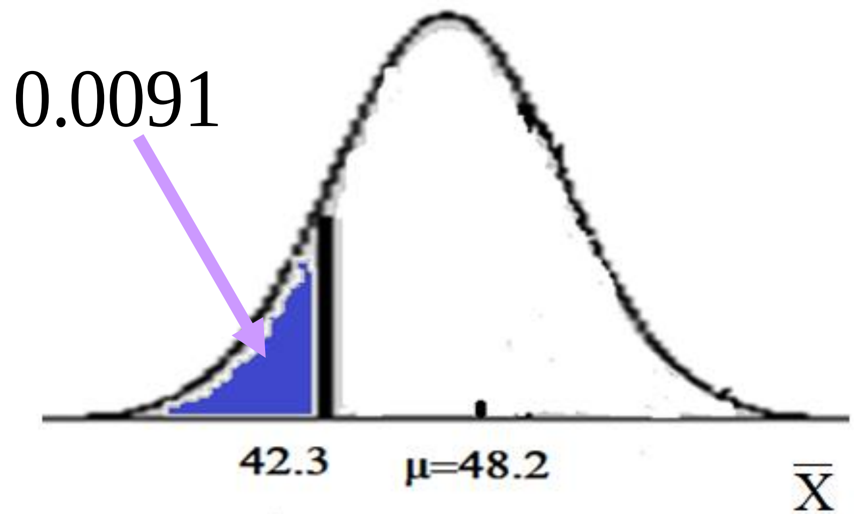
Sea **X** = "*eficacia de un procesador (tiempo de CPU) para realizar ciertas tareas*", (seg.), Con media μ **= 48.2 seg**. y varianza $\sigma^2 = 15^2$ **seg**^{2}.

Hay que hallar la probabilidad de que el contratista adquiera los procesadores.

La empresa decide **comprar** los procesadores **sólo si** en una muestra aleatoria de **36** de éstos, **el tiempo de CPU promedio** para realizar ciertas tareas es de a lo sumo **42.3 seg.**

$\overline{X}$ ="*eficacia promedio de los 36 procesadores seleccionados al azar para realizar ciertas tareas*".

$$P(\overline{X} \leq 42.3) = ?$$



***n* = 36** $\geq$ 30. Por el **TCL, entonces**:

$$\overline{X} \sim N\left(\mu = 48.2,\ \sigma^2/n = 15^2/36\right)$$

$$P(\overline{X} \leq 42.3) = 0.0091$$

El tiempo **promedio que requieren 36 CPU** para realizar ciertas tareas sea de a lo sumo **42.3 seg., es muy poco probable. Con lo cual la** empresa **no** piensa **comprar** los procesadores.

Ejemplo 2

Una empresa de mensajería que opera en la ciudad tarda un promedio de 35 minutos en llevar un paquete, con una desviación estándar de 10 minutos. Si durante el día de hoy han repartido 100 paquetes,

a) ¿Cuál es la probabilidad de que el tiempo promedio de entrega de hoy esté entre 30 y 35 minutos?

b) ¿Cuál es la probabilidad de que, en total, para los 100 paquetes hayan estado más de 60 horas?

a) ¿Cuál es la probabilidad de que el tiempo promedio de entrega de hoy esté entre 30 y 35 minutos?

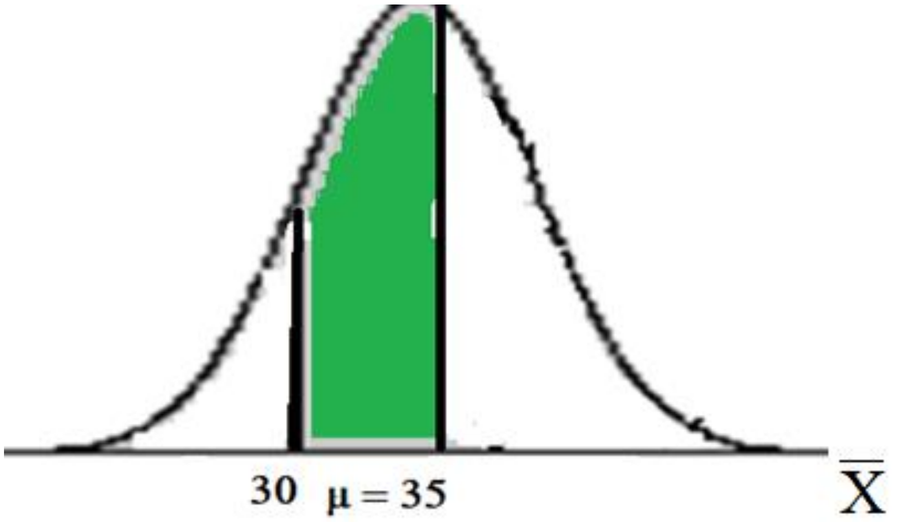
Sea la v.a. **X** = "*tiempo de entrega de un paquete* ", con media **μ = 35 min**. y varianza **σ^2 = 10^2 min^2**.

$\overline{X}$ ="*tiempo promedio de entrega de los 100 paquetes seleccionados al azar*".

Por el **Teorema Central del Límite**:

$\overline{X} \sim N(\mu = 35, \sigma^2/n = 10^2/100)$

La probabilidad de que el tiempo promedio de entrega de hoy esté entre 30 y 35 minutos equivale a calcular:



$$P(30 \leq \overline{X} \leq 35) =$$

$$= P(\overline{X} \leq 35) - P(\overline{X} \leq 30)$$

$$\cong 0.5 - 0 = \mathbf{0.5.}$$

b) ¿Cuál es la probabilidad de que, en total, para los 100 paquetes hayan estado más de 60 horas?

Sea la v.a. **X** = "tiempo de entrega de un paquete ",

con media μ **= 35 min**. y varianza σ^2 **= 10^2 min^2**.

P(tiempo total de entrega de los 100 paquetes > 60hs) =??

$\overline{X}$ ="tiempo promedio de entrega *de los 100 paquetes seleccionados al azar*".

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_{100}}{100} \Rightarrow 100 * \overline{X} = X_1 + X_2 + \cdots + X_{100}$$ **=Tiempo total**

$$\boxed{\overline{X} \sim N(\mu, \sigma^2/n).}$$

μ **= 35 min.** σ^2 **/n = 10^2 /100 min^2**

$$P(100 * \overline{X} > 60\,\text{hs.}) = ?$$

Como **n = 100** paquetes **≥ 30**, por el **TCL**

b) ¿Cuál es la probabilidad de que, en total, para los 100 paquetes hayan estado más de 60 horas?

$\overline{X}$="tiempo promedio de entrega *de los 100 paquetes seleccionados al azar*".

$$P(100^* \overline{X} > 60 \text{ hs.}) = ?$$

Como **n = 100** paquetes **≥ 30**, por el **TCL**

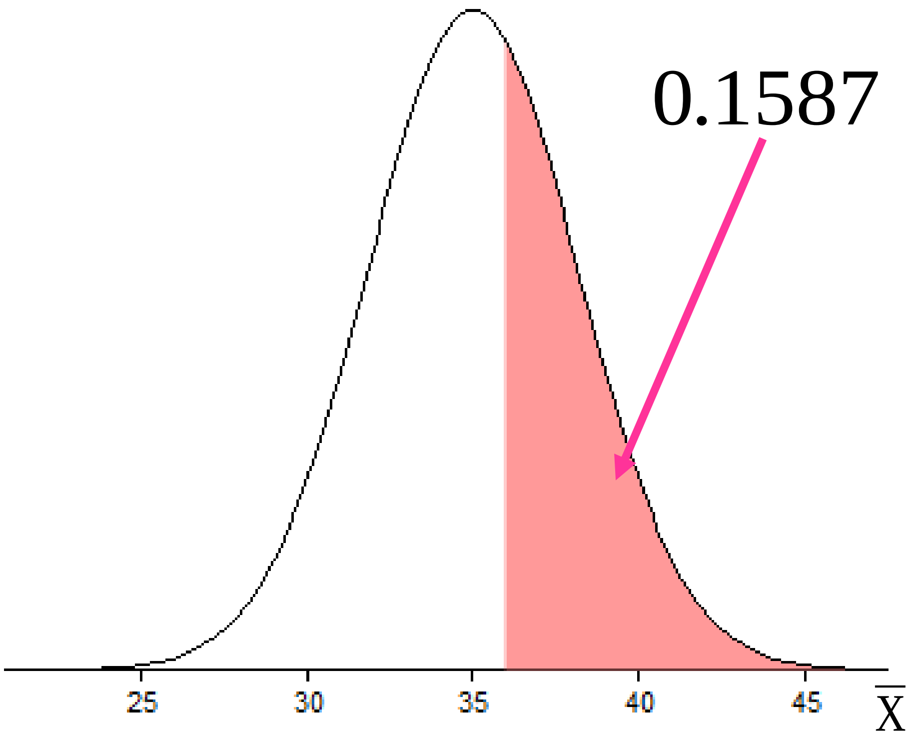
$$\overline{X} \sim N(\mu, \sigma^2/n).$$

$$P(100^* \overline{X} > 60 \text{ hs.}) =$$

$$= P(\overline{X} > 3600/100)$$

$$= P(\overline{X} > 36)$$

$$= 0.1587$$



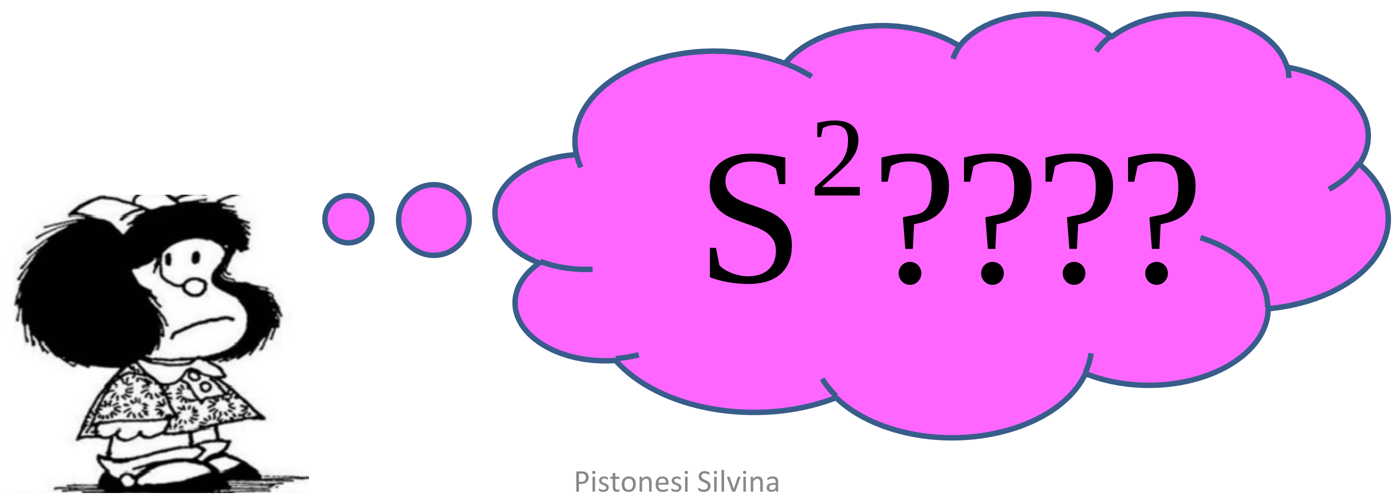
El **15.87 %** de las veces se insume más de 60 horas en repartir 100 paquetes.

Ejercicio

Se envían 70 mensajes independientes desde un centro de transmisión electrónica. Los mensajes son procesados secuencialmente, uno tras otro. El tiempo promedio de transmisión de cada mensaje es 0.2 minutos, con una varianza de 0.04 minutos^2. Encontrar la probabilidad de que los 70 mensajes se transmitan en menos de 12 minutos.

Estadístico:

Varianza muestral



Estadístico: Varianza muestral

Sea $X_1, X_2, \ldots, X_n$ una **muestra aleatoria** de tamaño ***n*** de la v.a. **X**

$$S^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1} = \frac{\left(\sum_{i=1}^{n} X_i^{\,2}\right) - n \, . \, \overline{X}^2}{n-1}$$

v.a. varianza muestral

Estadístico: Varianza muestral

$$S^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1}$$

Como es una **v.a**. es natural preguntarse:

1. ¿Cuál es su valor esperado? $E\left(S^2\right) = \sigma^2$

2. ¿Cuál será su distribución?

Sea $\mathbf{X_1, X_2, ..., X_{\mathit{n}}}$ una muestra aleatoria de tamaño ***n*** de la v.a. $\mathbf{X \sim N(\mu, \sigma^2)}$, entonces la v.a.

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

es una **v.a. Chicuadrado** con ν = ***n* - 1** grados de libertad.

Esta distribución interviene de manera especial al hacer **inferencias** respecto a la **varianza** de una distribución.

$$S^2 = \frac{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}{n-1} = \frac{\left(\sum_{i=1}^{n} Y_i^{\,2}\right) - n\,.\,\overline{Y}^2}{n-1}$$

Demostración de que: $E\left(S^2\right) = \sigma^2$ σ^2 = V(Y)

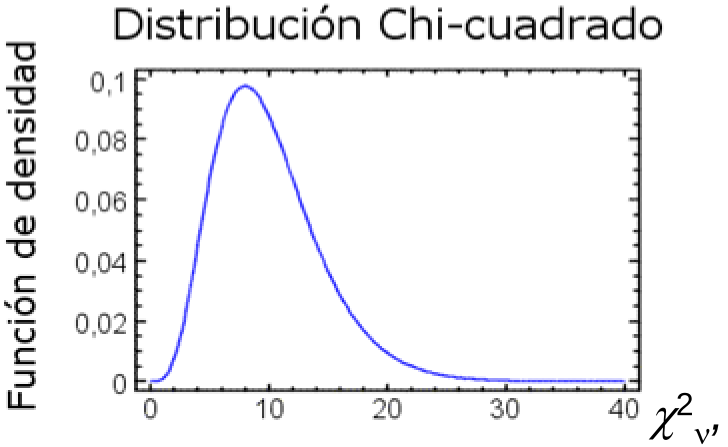
$$\begin{aligned}
\mathrm{E}[s^2] &= \mathrm{E}\left[\frac{1}{n-1}\sum_{i=1}^{n} Y_i^2 - \frac{n}{n-1}\overline{Y}^2\right] \\
&= \frac{1}{n-1}\left(\sum \mathrm{E}[Y_i^2] - n\,\mathrm{E}[\overline{Y}^2]\right) \\
&= \frac{1}{n-1}\left(n\,\mathrm{E}[Y_1^2] - n\,\mathrm{E}[\overline{Y}^2]\right) \\
&= \frac{n}{n-1}\left(\mathrm{Var}(Y_1) + \mathrm{E}[Y_1]^2 - \mathrm{Var}(\overline{Y}) - \mathrm{E}[\overline{Y}]^2\right) \\
&= \frac{n}{n-1}\left(\mathrm{Var}(Y_1) + \mu^2 - \frac{1}{n}\mathrm{Var}(Y_1) - \mu^2\right) \\
&= \frac{n}{n-1}\left(\frac{n-1}{n}\,\mathrm{Var}(Y_1)\right) \\
&= \mathrm{Var}(Y_1) \\
&= \sigma^2
\end{aligned}$$

Distribuciones Continuas usadas en Inferencia Estadística

Distribución Chi- cuadrado

La v.a. X se dice que tiene una ***distribución chi-cuadrado*** de parámetro ν, y se denota $\mathbf{X \sim \chi^2_\nu}$ si su función de densidad de probabilidad está dada por:

$$f(x) = \begin{cases} \dfrac{1}{2^{\nu/2}\,\Gamma(\frac{\nu}{2})}\, x^{\nu/2 - 1}\, e^{-x/2} & \text{si } x > 0 \\ 0 & \text{en el resto} \end{cases}$$



Características de la función de densidad:

•Esta caracterizada por un sólo parámetro ν, que recibe el nombre de *grados de libertad*.

•Definida sólo para valores positivos y cero de la va X.

•Es una curva asimétrica a derecha.

•Conforme $\nu \to \infty$ la distribución Chi-cuadrado se aproxima a la distribución Normal.

Ejemplo

Si tuviéramos, por ejemplo, la distribución de los salarios de los empleados de una Empresa dedicada a la fabricación de automóviles, en principio no podemos suponer la distribución normal ya la distribución es probablemente asimétrica con una cola hacia los salarios altos determinada por los salarios de los ejecutivos.

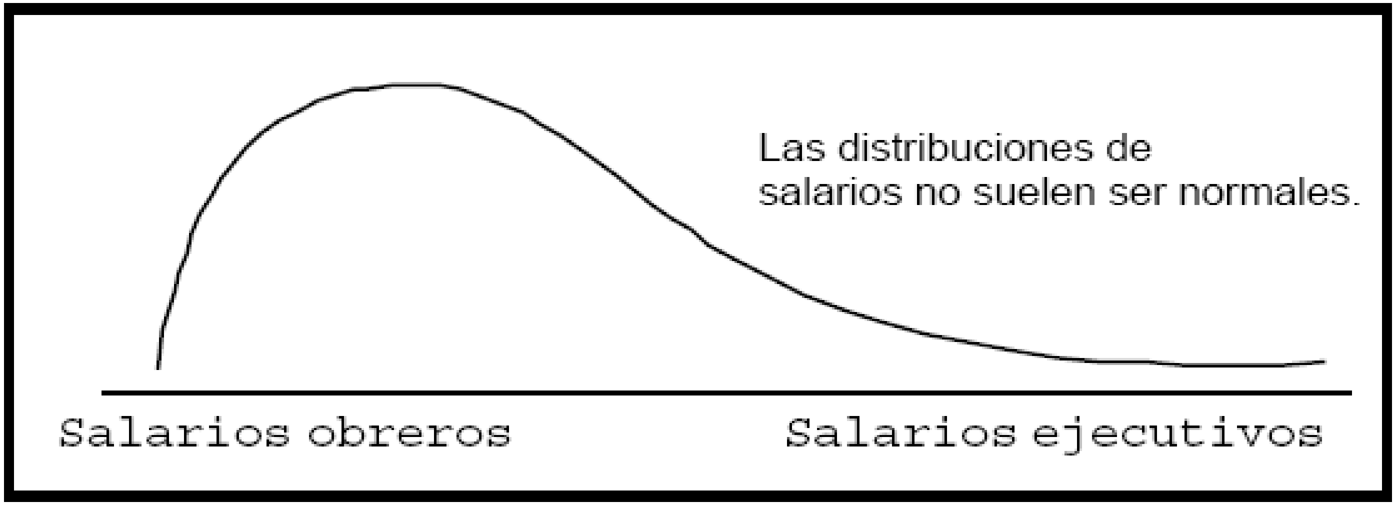
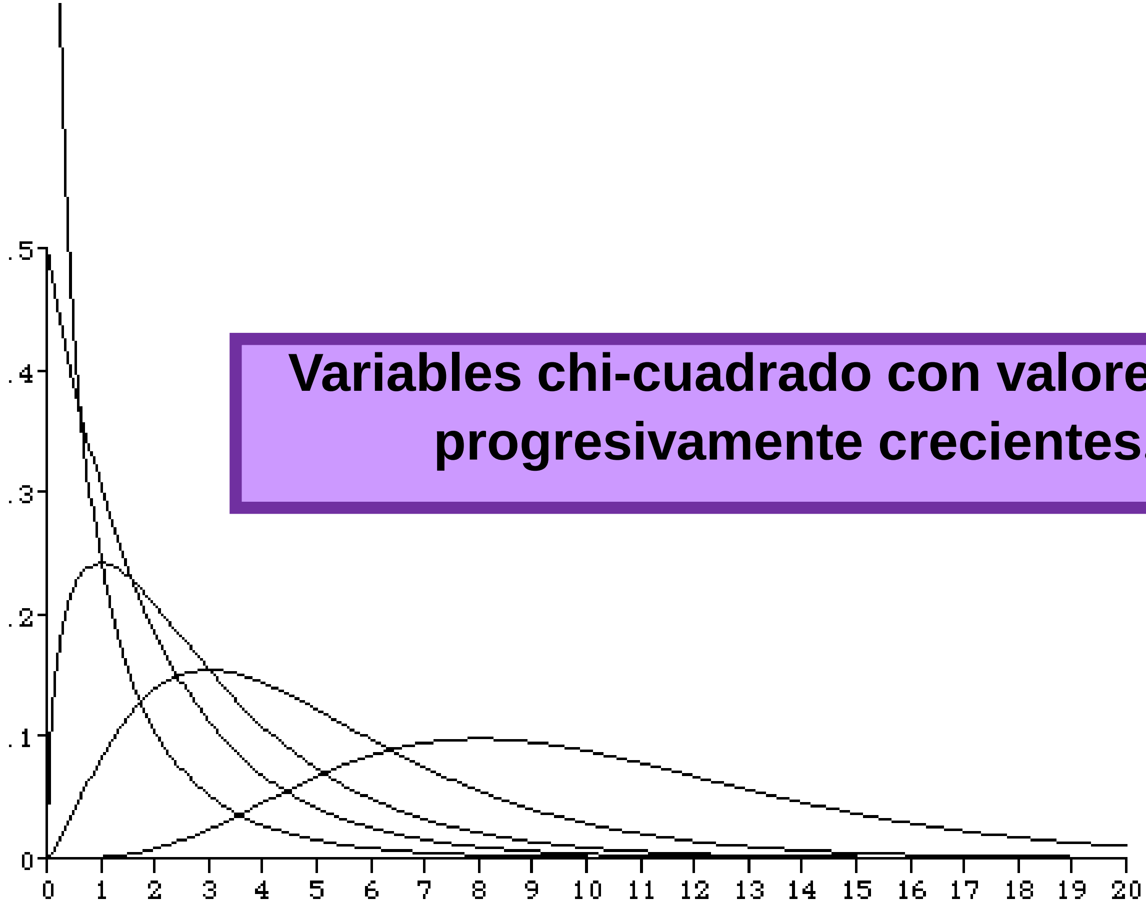


Figura 4: Distribución poblacional de los salarios de una empresa.

Valor Esperado y Varianza de una v.a. Chi-cuadrado

$$\boxed{E(\chi_\nu^2) = \nu} \quad \text{y} \quad \boxed{V(\chi_\nu^2) = 2\nu}$$



Variables chi-cuadrado con valores de ν progresivamente crecientes.

χ_ν^2

Manejo de la App

I) Hallar:

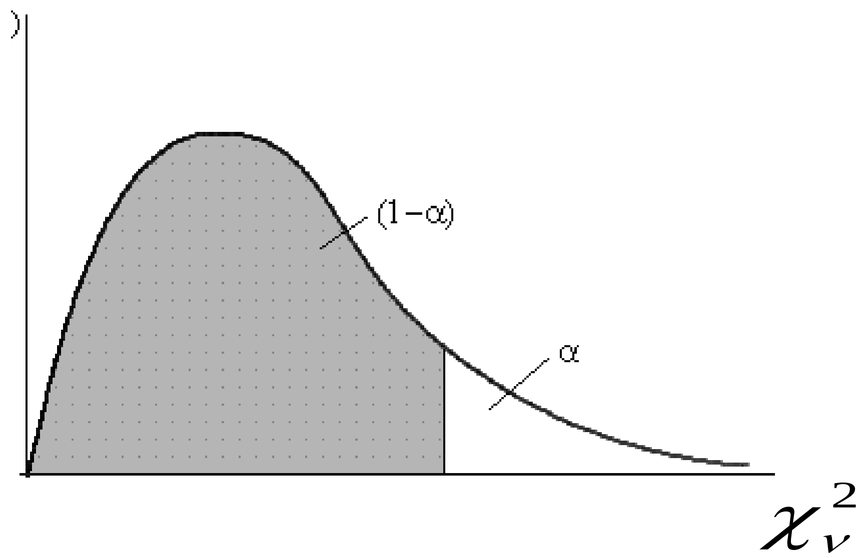
(a) $P(\chi^2_{30} < 18.49)$ (b) $P(\chi^2_{80} \geq 96.58)$ (c) $P(\chi^2_{3} > 9.35)$

II) Para la distribución Ji Cuadrado hallar el valor de $\chi^2_{\nu,\, 1-\alpha}$ de tal forma que:

(a) $P_{(}\chi^2_{6} < \chi^2_{6,\, 1-\alpha}) = 0.95$

(b) $P_{(}\chi^2_{21} > \chi^2_{21,\, 1-\alpha}) = 0.01$

(c) $P_{(}\chi^2_{10,\, 1-\alpha} < \chi^2_{10} < 23.21) = 0.95.$

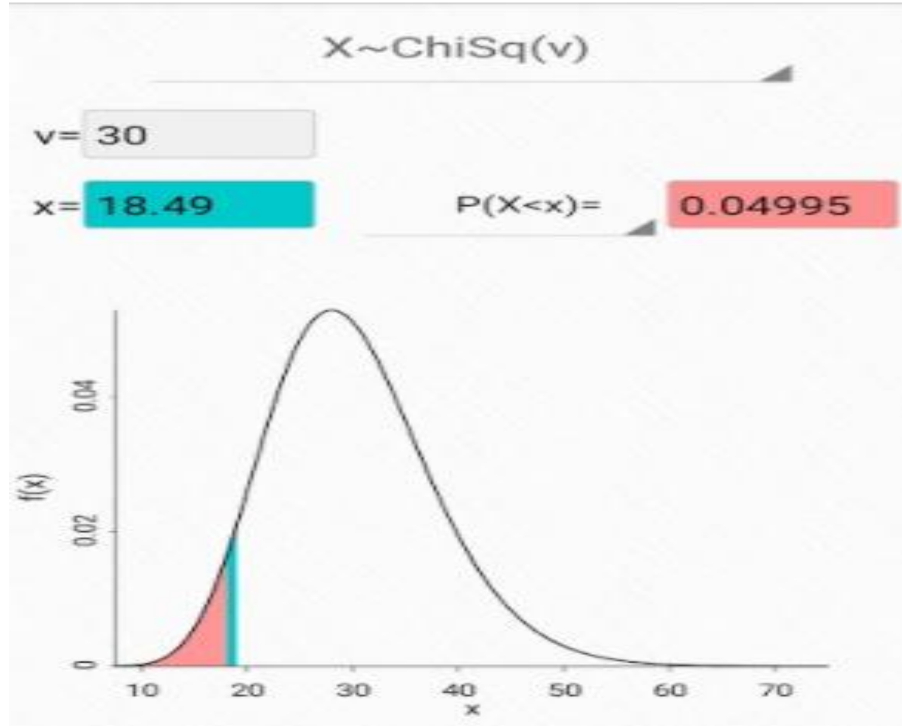


Manejo de la App

I) **Hallar:**

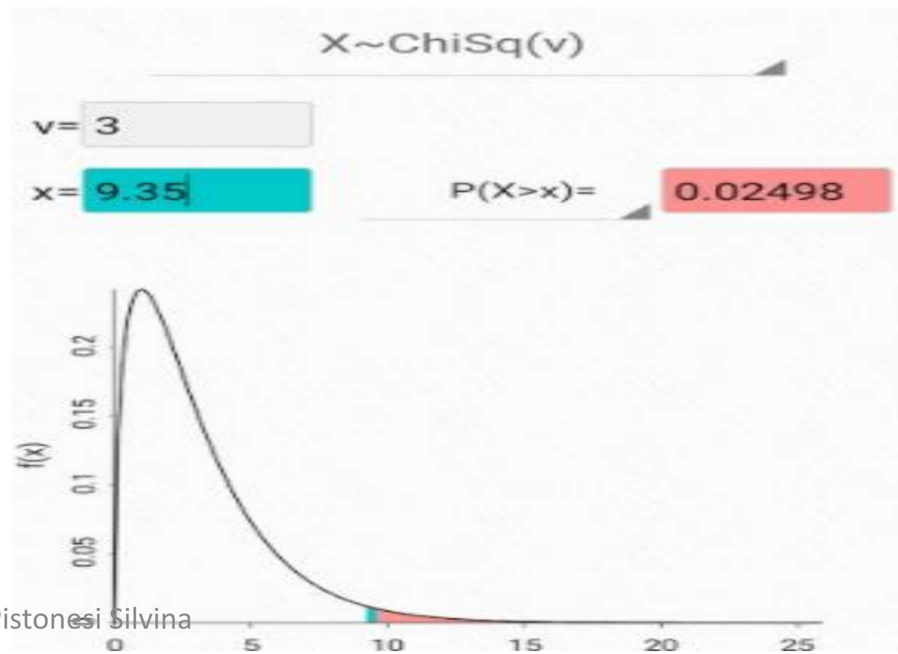
(a) P (χ^2_{30} < 18.49) =

= 0.04995



(c) P (χ^2_{3} > 9.35)

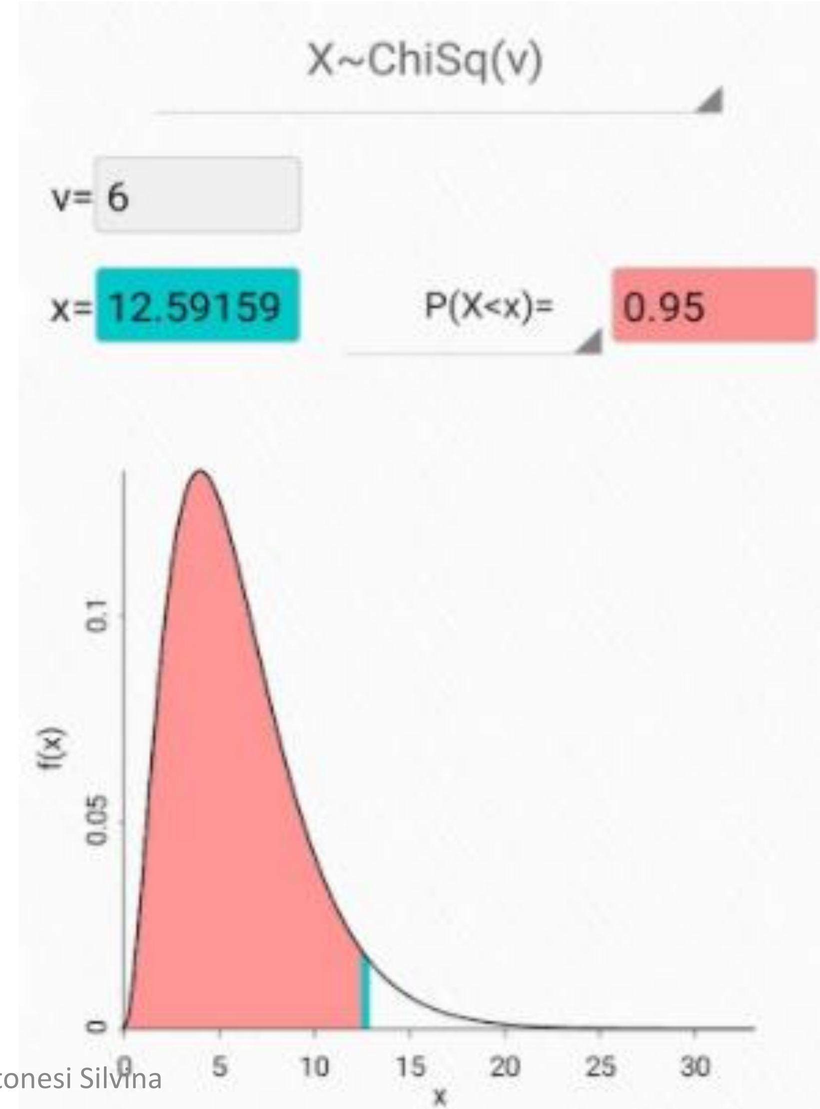
= 0.02498



Manejo de la App

II) Para la distribución Ji Cuadrado hallar el valor de $\chi^2_{\nu,\,1-\alpha}$ de tal forma que:

(a) $P\,(\chi^2_6 < \chi^2_{6,\,1-\alpha}) = 0.95$



Luego:

$\chi^2_{6,\,1-\alpha}$ **= 12.59159**

(c) $P(\chi^2_{10,\,1-\alpha} < \chi^2_{10} < 23.21) = 0.95.$

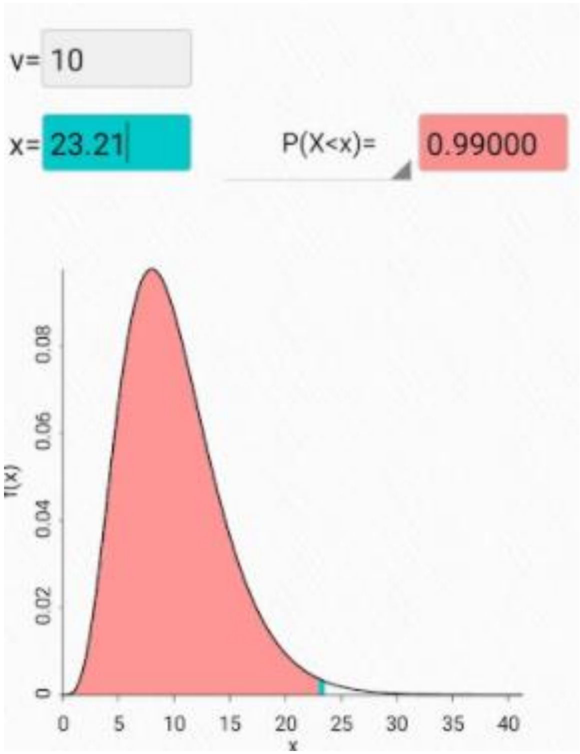
$0.95 = P(\chi^2_{10,\,1-\alpha} < \chi^2_{10} < 23.21) =$

$= P(\chi^2_{10} < 23.21) - P(\chi^2_{10} < \chi^2_{10,\,1-\alpha})$

$0.95 = 0.\ \ 0.99\ \ - P(\chi^2_{10} < \chi^2_{10,\,1-\alpha})$

$P(\chi^2_{10} < \chi^2_{10,\,1-\alpha}) = 0.04$

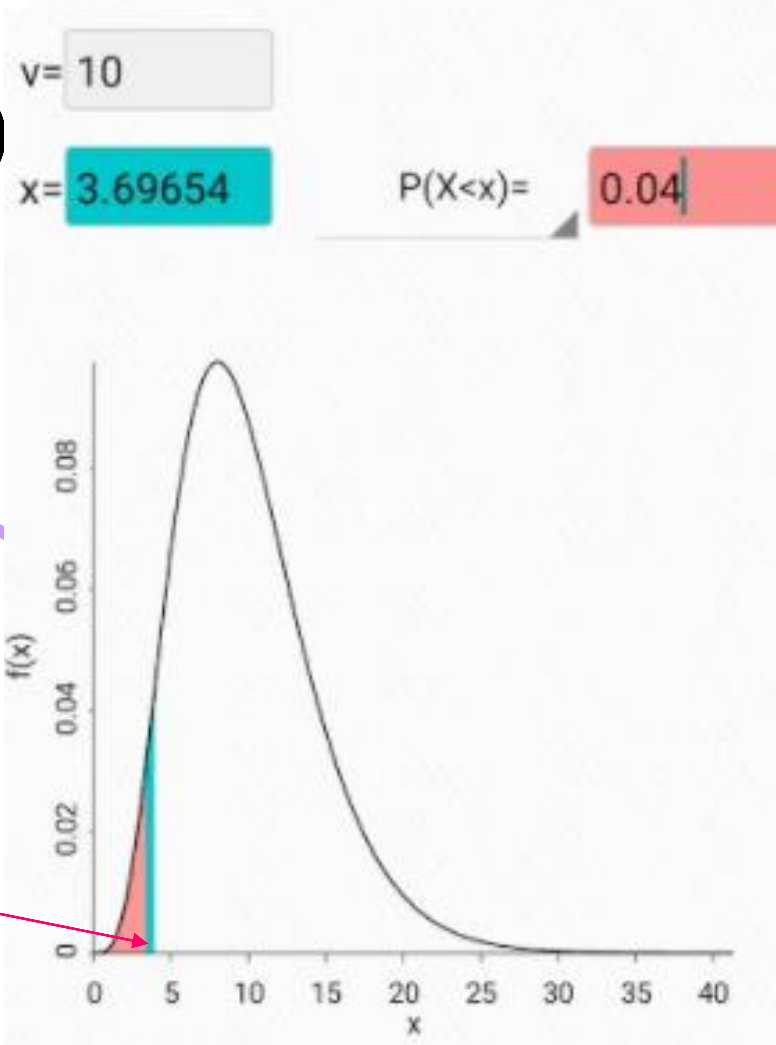
Luego $\chi^2_{10,\,1-\alpha} =$



(c) $P(\chi^2_{10,\,1-\alpha} < \chi^2_{10} < 23.21) = 0.95.$

$$= P(\chi^2_{10} < 23.21) - P(\chi^2_{10} < \chi^2_{10,\,1-\alpha})$$

$$0.95 = \boxed{0.99} - P(\chi^2_{10} < \chi^2_{10,\,1-\alpha})$$



$$P(\chi^2_{10} < \chi^2_{10,\,1-\alpha}) = 0.04$$

App

Luego

$$\chi^2_{10,\,1-\alpha} = 3.69654$$

Ejemplo

La autoridad sanitaria de un país decide llevar a cabo una investigación sobre los residuos que producen las empresas de un determinado sector anualmente. Seleccionada una muestra aleatoria simple de 9 empresas y suponiendo que la cantidad de residuos producidos por una empresa por año se distribuye normalmente con media 23 Tn y desviación estándar de 6 Tn, calcular la **probabilidad** de que la **varianza** de la **cantidad de residuos** que producen las empresas **muestreadas** sea superior a 60.12Tn^2.

X = "*cantidad de residuos que produce una empresa por año* ", **(Tn)**

X ~ N(μ = 23, σ^2 = 6^2), se tomó una muestra aleatoria de **9** empresas

P(S^2 > 60.12) = ???

$$\frac{(n-1)S^2}{\sigma^2} = \chi_{n-1}^2 \quad \Rightarrow \quad S^2 = \frac{\sigma^2 \chi_{n-1}^2}{(n-1)}$$

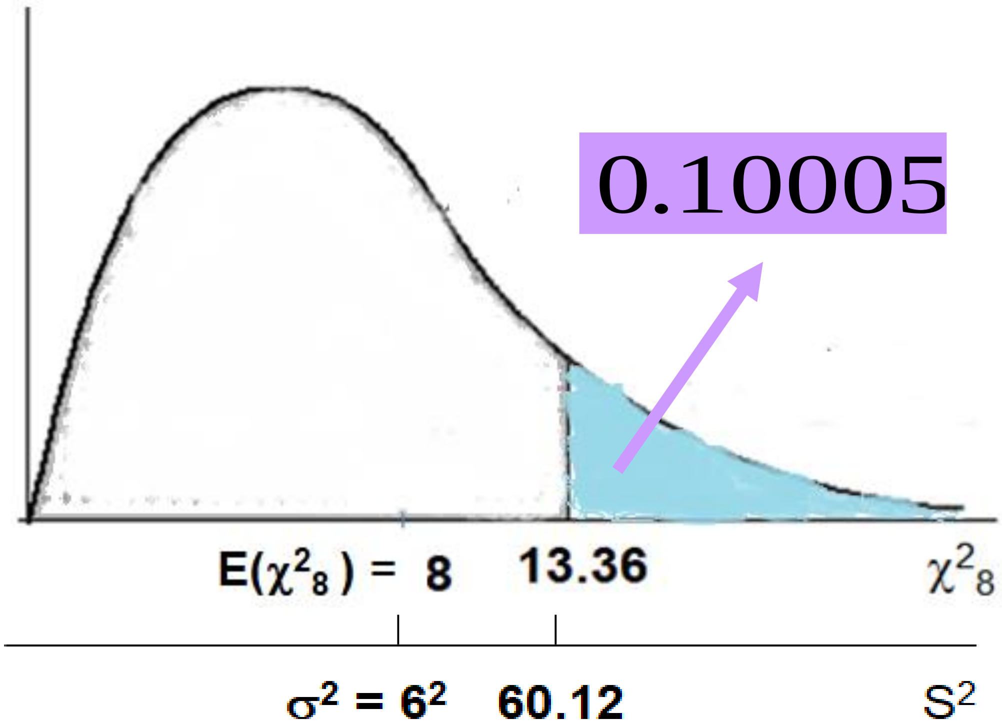
$$P(S^2 > 60.12) = P(\frac{\chi_8^2 6^2}{9-1} > 60.12) = ?$$

$$P(S^2 > 60.12) = P\left(\frac{\chi_8^{\,2}\, 6^2}{9-1} > 60.12\right) = ?$$

$$= P\left(\chi_8^{\,2} > \frac{60.12 \bullet 8}{36}\right)$$

$$= P(\chi_8^{\,2} > 13.36) =$$

$$= 0.10005$$



En el **10.005%** de las veces que se seleccionan 9 empresas al azar la varianza de la cantidad de residuos que producen es superior a 60.12Tn^2.

Ejercicio:

El voltaje de salida de la fuente de alimentación de una computadora de cierta marca sigue una distribución normal con media de 5 V y desvío estándar 0.60 V.

a) ¿Cuál es la probabilidad de que el voltaje promedio de una muestra de 8 fuentes de alimentación se aleje de su valor esperado en más de 0.30 V?

b) Si se eligen 16 fuentes al azar, ¿cuál es la probabilidad de que la varianza del voltaje de esa muestra sea superior a 0.4635 V^2?