Estadística
T6. Muestreo y Teorema Central del Límite

José Rodellar - Yadira Salazar – Kevin Barrera

Escola d'Enginyeria de Barcelona Est (EEBE)
Departament de Matemàtiques
Universitat Politècnica de Catalunya (UPC)

Curso 2025 - 26

- Teorema Central del Límite
- Introducción a la inferencia
- Muestreo
- Media y varianza de la muestra
- Problemas

Introducción a la inferencia

Población

- Experimento aleatorio
- Variable aleatoria X
- Modelo de probabilidad f(x)

  $E(X) = \mu$

  $Var(X) = \sigma^2$

MUESTREO

Muestra aleatoria

$X_1, X_2, \cdots, X_n$

INFERENCIA

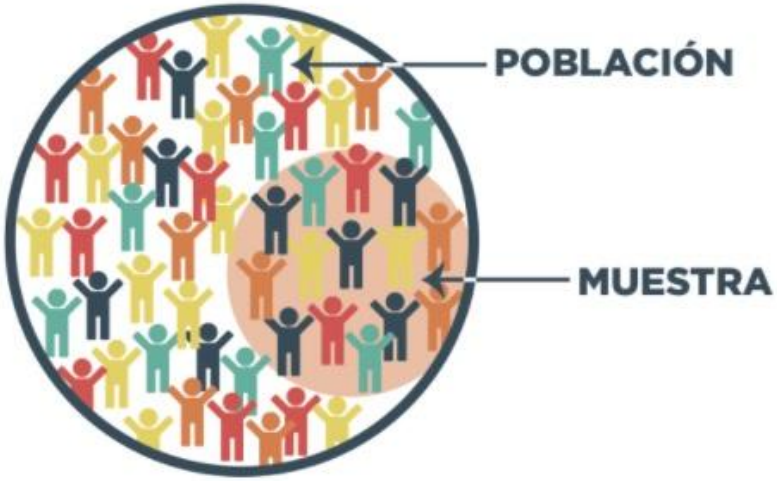
Modelo de probabilidad para toda la población

VISTO HASTA AHORA

A PARTIR DE AHORA

Introducción a la inferencia

En la mayoría de los casos, no podemos examinar toda una población debido al tiempo, costo o esfuerzo que implicaría. En cambio, tomamos una **muestra** y tratamos de inferir algo sobre la **población** general.



La inferencia estadística es el proceso de usar datos de una muestra para hacer generalizaciones sobre una población más grande.

Teorema central del límite

INTERPRETACIÓN

El Teorema Central del Límite (TCL) es uno de los principios fundamentales en estadística. Lo que nos dice es bastante **sorprendente**: incluso si una población no se distribuye normalmente, la **distribución de las medias** de muchas **muestras tomadas de esa población** se acercará a una **distribución normal** (también conocida como distribución gaussiana o campana de Gauss). Esto ocurre siempre y cuando el tamaño de las muestras sea suficientemente grande (**una regla general es que el tamaño sea mayor o igual a 30**).

Teorema central del límite

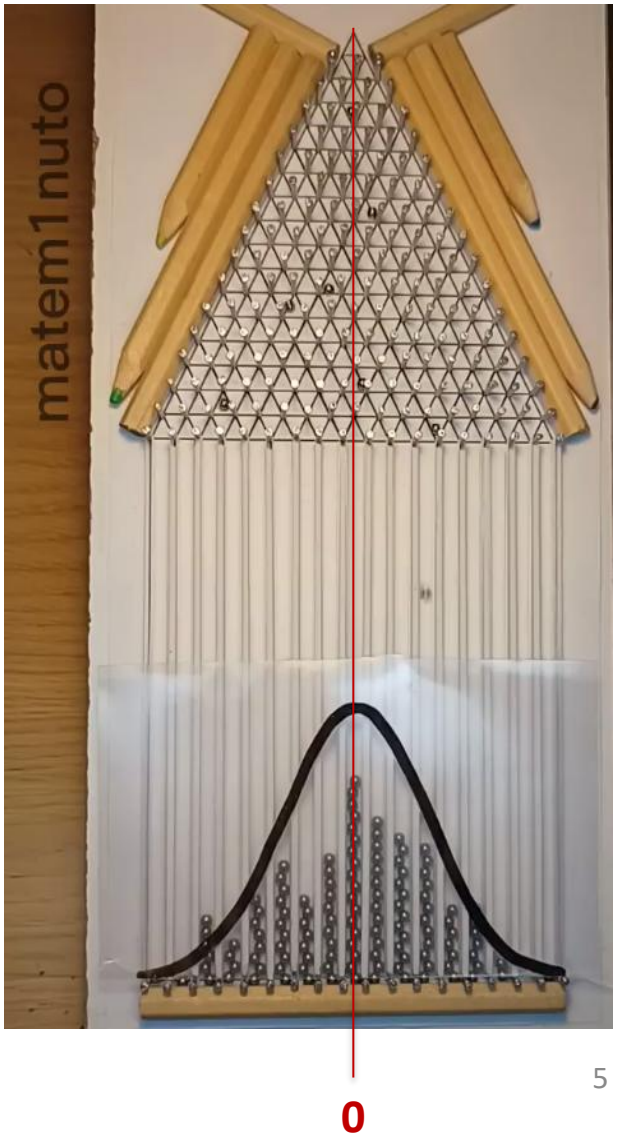
Ejemplo: Experimento de Galton

Un tablero vertical con filas de clavos intercalados. Se dejan caer bolas que van rebotando aleatoriamente hacia la derecha o la izquierda al chocar con los clavos.

Finalmente, las bolas se acumulan en unas guías y pueden contarse.

En este ejemplo hay 16 filas de clavos y por tanto hay 17 guías. Por la simetría, podemos considerar la guía central como cero y el resto de la siguiente forma:

| -8 | -7 | · · · |-1 | **0** | 1 | · · · | 7 | 8 |

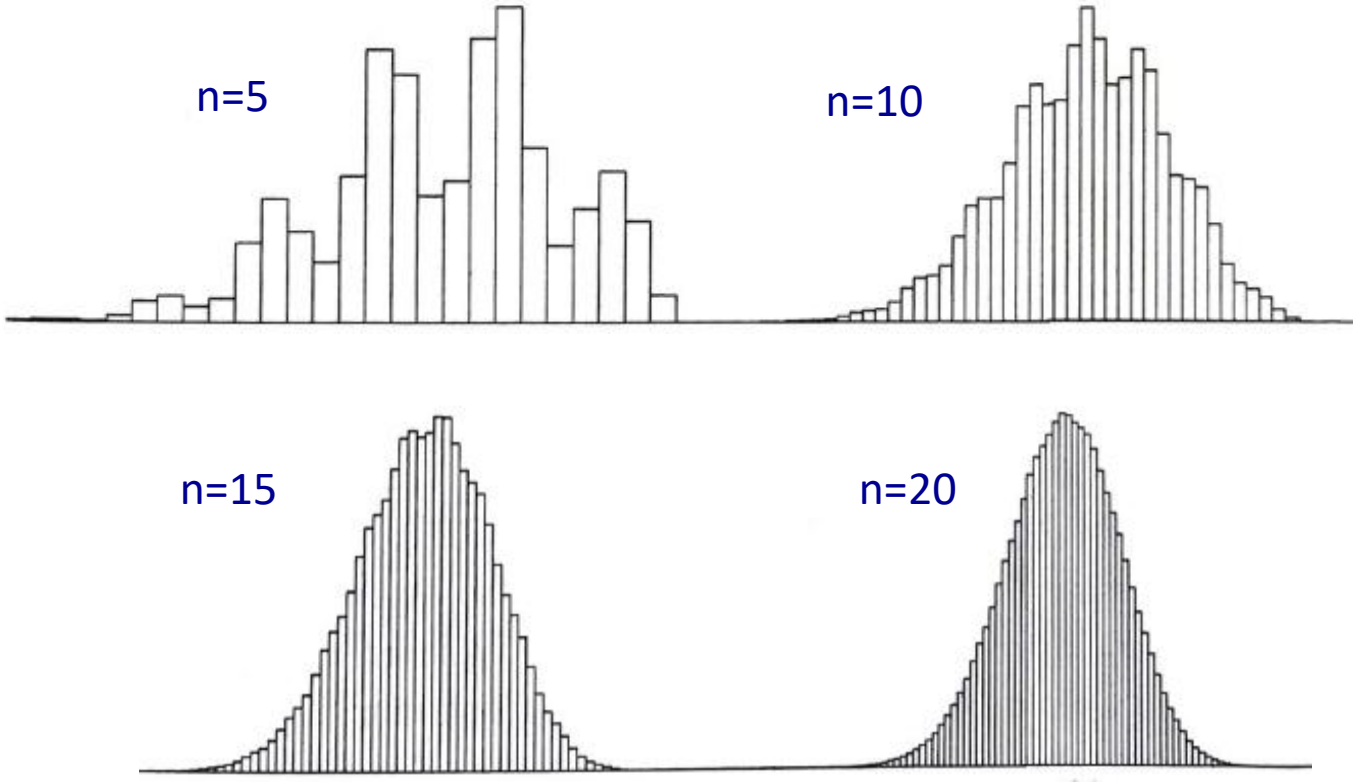


matem1nuto

Teorema central del límite

Ejemplo

$p_1 = 0.2; p_2 = 0.1; p_3 = p_4 = 0; p_5 = 0.3; p_6 = 0.4$



Cuanto más asimétrica es la distribución de las variables **Xi** alrededor de su valor medio, mayor tiene que ser el valor de n para alcanzar una buena aproximación con la normal.

Muestra aleatoria simple

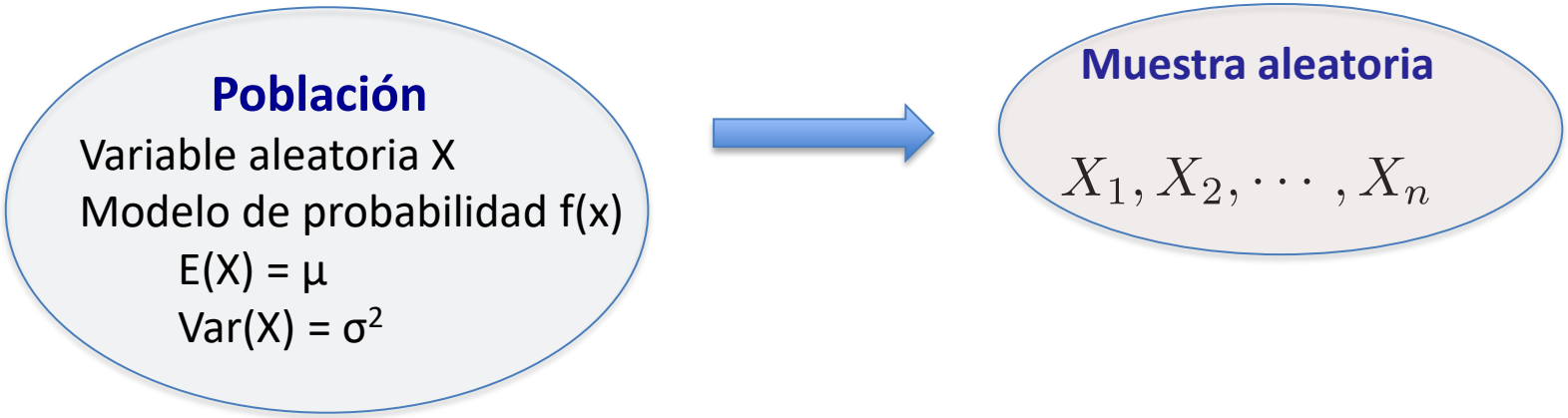
Una vez que tenemos esta muestra, **¿cómo podemos usarla para hacer inferencias sobre la población completa?**

Nos centraremos en tres conceptos estadísticos clave que sirven como **estimadores** de parámetros poblacionales:

- **Suma Muestral**: Nos permite estimar la suma total de la población.
- **Media Muestral:** Proporciona un estimado de la media poblacional.
- **Varianza Muestral:** Ofrece una estimación de la dispersión en la población

Estos estimadores son la base para la inferencia estadística.

Suma Muestral



La **suma muestral** es la suma total de todos los valores observados en una muestra. Para una muestra aleatoria simple X_1, X_2, …, X_n, se define como:

$$S = \sum_{i=1}^{n} X_i$$

Suma Muestral

Supón que estás fabricando piezas mecánicas.

- Cada pieza tiene un error de longitud: Xi
- El error total en un lote de 50 piezas:

$$S_{50=}X_1 + X_2 + X_3 + X_5 + X_{50}$$

¿Para qué sirve?

- Ver si el error acumulado supera un límite
- Evaluar si el proceso es aceptable

Un sistema mide consumo eléctrico cada hora.

- Xi: consumo en la hora i
- Energía total diaria:

$$S_{24=}X_1 + X_2 + X_3 + X_5 + X_{24}$$

Aplicación:

- Calcular consumo total
- Optimizar eficiencia energética

Suma Muestral: Propiedades

Si cada observación viene de una distribución con media μ y varianza σ^2 entonces:

$$E[S] = n \cdot \mu$$

$$\mathrm{Var}(S) = n \cdot \sigma^2$$

Ejemplo: Suma Muestral

El número de horas que dedican a estudiar los alumnos de una universidad se distribuye normalmente con una media de 15 horas y una varianza de 9 horas.

(a) ¿Cuál es la probabilidad de que la suma total de horas de estudio en un grupo de 10 estudiantes seleccionados de forma aleatoria sea mayor a 160 horas?

(b) ¿Cuál es la probabilidad de que la suma muestral de horas de estudio en una muestra de 10 estudiantes sea menor a 130 horas?

(c) Si seleccionamos varias muestras de 10 estudiantes cada una, ¿cuál es el valor esperado y la varianza de la suma muestral total de horas de estudio en cada muestra?

(d) ¿Cuál es la probabilidad de que un grupo de 10 estudiantes tenga un promedio de horas de estudio entre 140 y 160 horas?

Teorema central del límite para la suma muestral

Sea una **muestra aleatoria simple** $X_1, X_2, \cdots, X_n$

de una variable aleatoria con **valor medio y desviación típica** $\mu,\ \sigma$

Consideremos la **suma muestral**: $S_n = X_1 + X_2 + \cdots + X_n$

Su **valor esperado** y su **varianza** son

$$E(S_n) = n\mu$$

$$Var(S_n) = n\sigma^2$$

Según el TCL, se cumple que

$$\lim_{n\to\infty} \frac{S_n - n\mu}{\sigma\sqrt{n}} \Rightarrow N(0,1)$$

Media Muestral

Población

Variable aleatoria X

Modelo de probabilidad f(x)

$E(X) = \mu$

$Var(X) = \sigma^2$

Muestra aleatoria

$$X_1, X_2, \cdots, X_n$$

Calculamos la **media muestral**

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

La **media muestral** es el promedio de los valores observados en una muestra. Se calcula como:

$$\bar{X} = \frac{1}{n}S = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Media Muestral: Propiedades

La media muestral es un estimador insesgado de la media poblacional μ. Es decir, la media muestral es un buen estimador de la media poblacional porque su **valor esperado** es igual a la **media poblacional**

$$E[\bar{X}] = \mu$$

$$\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

Teorema central del límite para la media muestral

Sea una **muestra aleatoria simple** $X_1, X_2, \cdots, X_n$

de una variable aleatoria con **valor medio y desviación típica** $\mu,\ \sigma$

Consideremos la **media muestral**: $\bar{X} = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$

Su **valor esperado** y su **varianza** son $\begin{cases} E(\bar{X}) = \mu \\ Var(\bar{X}) = \sigma^2_{\bar{X}} = \dfrac{\sigma^2}{n} \end{cases}$

Según el TCL, se cumple que

$$\lim_{n\to\infty} \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \hookrightarrow N(0,1)$$

Teorema central del límite para la media muestral

$$\lim_{n\to\infty} \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \hookrightarrow N(0,1)$$

Cuando n tiende a infinito, la media muestral tiende a tener una distribución normal con

$$\text{Valor medio } \mu$$

$$\text{Desviación estándar } \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

Esto es independiente de cuál sea la distribución de la variable X en el seno de la población.

Gracias al TCL, sabemos que **la media de una muestra aleatoria** suficientemente grande estará muy próxima al valor medio verdadero de toda la población. Además la dispersión se reducirá al aumentar el tamaño de la muestra.

Media Muestral: Propiedades

Conclusión

$$E(\bar{X}) = \mu$$
$$Var(\bar{X}) = \sigma^2_{\bar{X}} = \frac{\sigma^2}{n}$$

El **valor esperado** de la media muestral coincide con el de toda la población.

La **varianza** de la media muestral es 1/n veces la de la población.

Cuanto mayor sea n, menor será la dispersión en la muestra.

Este resultado es independiente de cuál es el modelo de probabilidad de la variable X en la población.

Ejemplo: Media Muestral

El tiempo que los empleados de una empresa dedican a actividades laborales diariamente sigue una distribución normal con una media de 6 horas y una varianza de 1.5 horas.

(a) ¿Cuál es la probabilidad de que el promedio de horas laborales diarias en una muestra aleatoria de 25 empleados sea menor a 5.8 horas?

(b) ¿Cuál es la probabilidad de que la media de horas laborales diarias en una muestra de 25 empleados se encuentre entre 5.5 y 6.5 horas?

(c) Si seleccionamos múltiples muestras de 25 empleados cada una, ¿cuál es el valor esperado y la varianza del promedio de horas laborales diarias en cada muestra?

Varianza Muestral

Población

Variable aleatoria X

Modelo de probabilidad f(x)

$E(X) = \mu$

$Var(X) = \sigma^2$

Muestra aleatoria

$$X_1, X_2, \cdots, X_n$$

La **varianza muestral** mide la dispersión de los datos en una muestra. Se define como:

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

Varianza Muestral

Ejemplo del cálculo de la varianza muestral:

Datos: 1,3,5
Media: 3

$$\boldsymbol{S^2_{n-1}} = \frac{(1-3)^2 + (3-3)^2 + (5-3)^2}{3-1} = \mathbf{4}$$

Varianza Muestral: Propiedades

La varianza muestral S^2_{n-1} es un estimador insesgado de la varianza poblacional σ^2 . Es decir, el valor esperado de la varianza muestral coincide con la varianza poblacional σ^2

$$E[S^2] = \sigma^2$$

Varianza Muestral: Propiedades

Población: Supongamos que **nuestra población** completa es

$$\{1,2,3,4,5\}$$

Calculamos la **varianza poblacional (σ^2)**

1.- Media Poblacional

$$\mu=(1+2+3+4+5)/5 = \mathbf{3}$$

2.- Varianza Poblacional

$$\frac{(1-3)^2+(2-3)^2+(3-3)^2+(4-3)^2+(5-3)^2}{5} = 2$$

Varianza Muestral: Propiedades

Muestra	Media muestral $\overline{X}$	Cálculo $S^2 = \frac{\sum(X_i-\overline{X})^2}{n-1}$	Varianza
(1, 2, 3)	2	$\frac{(1-2)^2+(2-2)^2+(3-2)^2}{2} = \frac{1+0+1}{2}$	1
(2, 3, 4)	3	$\frac{(2-3)^2+(3-3)^2+(4-3)^2}{2} = \frac{1+0+1}{2}$	1
(3, 4, 5)	4	$\frac{(3-4)^2+(4-4)^2+(5-4)^2}{2} = \frac{1+0+1}{2}$	1
(1, 3, 5)	3	$\frac{(1-3)^2+(3-3)^2+(5-3)^2}{2} = \frac{4+0+4}{2}$	4
(1, 2, 5)	2.67	$\frac{(1-2.67)^2+(2-2.67)^2+(5-2.67)^2}{2} = \frac{2.78+0.44+5.44}{2}$	4.33
(2, 4, 5)	3.67	$\frac{(2-3.67)^2+(4-3.67)^2+(5-3.67)^2}{2} = \frac{2.78+0.11+1.78}{2}$	2.33

$$E[S^2] = \frac{1+1+1+4+4.33+2.33}{6} = \frac{13.66}{6} \approx 2.28$$

Varianza Muestral: Propiedades

La varianza de una varianza muestral (Var S^2_{n-1}) qué tan variable es la varianza muestral cuando se extraen diferentes muestras de una misma población.

$$Var(S^2) = \frac{2\sigma^4}{n-1}$$

Varianza Muestral: Distribución Chi-cuadrado

- Para el cálculo de probabilidades relacionadas con la varianza muestral, ya no se utiliza la distribución normal, sino la distribución chi-cuadrado (χ^2).

- Esto se debe a que, al calcular la varianza, trabajamos con valores elevados al cuadrado (las desviaciones respecto a la media).

- En este caso, los datos se vuelven cuadráticos y siempre positivos, y su comportamiento se describe mejor mediante la distribución chi-cuadrado, que surge precisamente al sumar cuadrados de datos de origen normal.

Varianza Muestral: Distribución Chi-cuadrado

La varianza muestral S^2 tiene una relación con la distribución Chi-cuadrado (χ^2) cuando la población de origen es **normal**.

El estadístico $(n-1)\frac{S^2}{\sigma^2}$ sigue una distribución

Chi-cuadrado con ***n-1*** grados de libertad si la distribución es normal:

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

Varianza Muestral: Distribución Chi-cuadrado

- Este valor representa qué tan grande o pequeña es la varianza muestral en comparación con la varianza poblacional, pero expresado en la escala de la distribución chi-cuadrado.

- En otras palabras, es una forma de "tipificar" la varianza muestral, igual que cuando usamos un valor Z para las medias, pero aquí usamos χ^2 porque trabajamos con valores al cuadrado.

- Este valor de χ^2 nos servirá luego para el cálculo de probabilidades, ya que permite ubicar la varianza muestral dentro de la curva chi-cuadrado y determinar qué tan probable o rara es la dispersión observada en la muestra.

RESUMEN

	Suma muestral	Media muestral	Varianza muestral
Ecuación	$S = \sum_{i=1}^{n} x_i$	$\bar{X} = \frac{1}{n} \cdot S = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$	$S^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{X})^2$
E ()	$E(S) = n \cdot \mu = \mu_s$	$E(\bar{X}) = \mu = \mu_{\bar{X}}$	$E(S^2) = \sigma^2 = \mu_{s^2}$
Var ()	$Var(S) = n \cdot \sigma^2 = \sigma_s^2$	$Var(\bar{X}) = \frac{\sigma^2}{n} = \sigma_{\bar{X}}^2$	$Var(S^2)$ $X^2{}_{n-1} = (n-1)\frac{S^2}{\sigma^2}$

Ejemplo: Varianza Muestral

Un fabricante de componentes electrónicos desea controlar la consistencia en la duración de las baterías que produce. Se sabe que la duración de estas baterías sigue una distribución normal con una media de 20 unidades y una varianza de 16 unidades.

(a) ¿Cuál es la probabilidad de que la varianza muestral de la duración de las baterías en una muestra de 30 baterías sea menor a 25 unidades?

(b) ¿Cuál es la probabilidad de que la varianza muestral de la duración de las baterías en una muestra de 30 baterías sea mayor a 18 unidades?

(c) ¿Cuál es la probabilidad de que la varianza muestral de la duración de las baterías en una muestra de 30 baterías esté entre 10 y 20 unidades?

EJERCICIO

Una biblioteca está investigando el comportamiento de sus usuarios respecto al número de libros que leen al mes. El número de libros leídos por los usuarios sigue una distribución normal con una media de 30 libros y una desviación estándar de 5 libros. Contesta las siguientes preguntas:

(a) Si se seleccionan 18 usuarios al azar, ¿cuál es la probabilidad de que la suma total de libros leídos por el grupo sea mayor que 550 libros durante un mes?

(b) Si se seleccionan 50 usuarios al azar, ¿cuál es la probabilidad de que el promedio de libros leídos por el grupo sea menor que 28 libros durante un mes?

(c) Si se selecciona un usuario al azar, ¿cuál es la probabilidad de que lea al menos 25 libros en un mes?

(d) Si se selecciona un grupo de 20 usuarios al azar, ¿cuál es la probabilidad de que el promedio de libros leídos por el grupo en conjunto sea mayor que 31 libros durante un mes?

(e) Si se selecciona un grupo de 25 usuarios al azar, ¿cuál es la probabilidad de que la varianza del número de libros leídos por el grupo sea menor que 12 libros?