

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO ENES MÉRIDA LICENCIATURA EN ECOLOGÍA

INTRODUCCIÓN A LA ESTADÍSTICA Tema 2. Medidas de posición y dispersión

Prof. Edlin J. Guerra Castro

El TRUCO en el diseño de experimentos es construir una experiencia en que se ponga a prueba la hipótesis nula.

¿cuáles son los elementos de una hipótesis nula?

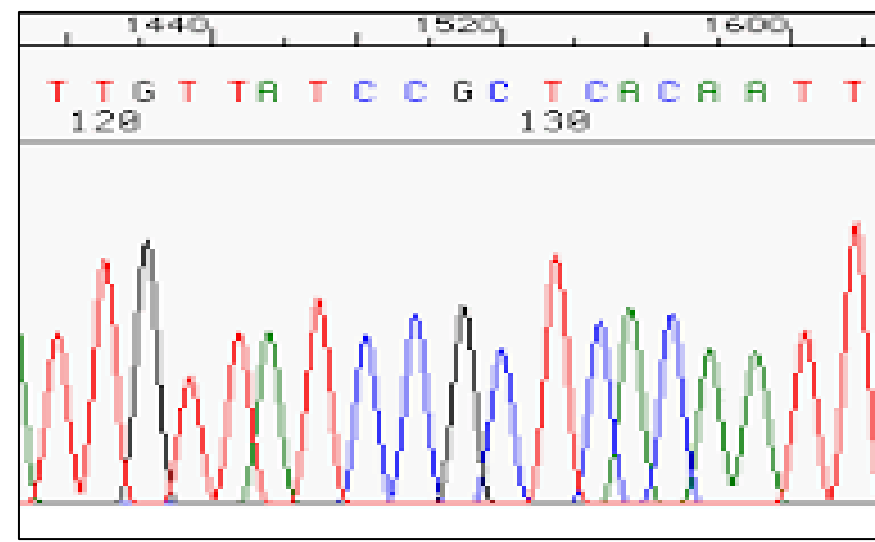
- variable causal
- variable respuesta

Variables en ciencias naturales

¿qué es un variable?

- Es una propiedad en la que cada elemento de una población difiere en cierto modo.

Sokal y Rohlf (1995)
Quinn & Keough (2002)



Cuánto observamos? A quiénes o qué?

Población target vs. Población estadística

Todas las
entidades
sobre las
cuáles
extraemos
información.

Todas las
expresiones de
una variable
sobre las cual se
quiere hacer
inferencia. Tal
variable ayuda a
describir una
población
biológica

Tipos de variables

Continuas.- Son variables que pueden asumir cualquier valor imaginable dentro de un intervalo no numerable de valores (se miden); ej.

Discretas.- Son variables que solo pueden asumir ciertos valores en un intervalo, por lo general enteros (se cuentan); ej.

Rango.- Variables medidas en una escala donde las unidades no mantienen una relación constante o conocida entre si pero que se pueden organizar de una manera jerárquica (se ordenan); ej.

Atributo.- Variables medidas en una escala donde las unidades no mantienen una relación constante ni conocida, no se pueden organizar en forma jerárquica (se clasifican); ej.

Derivadas.- Son variables que provienen de la relación matemática de otras variables inicialmente medidas; ej.

Transformadas.- Son variables a las cuales se les aplica una operación matemática con la idea de modificar la escala en la que han sido medidas; ej.



elemento

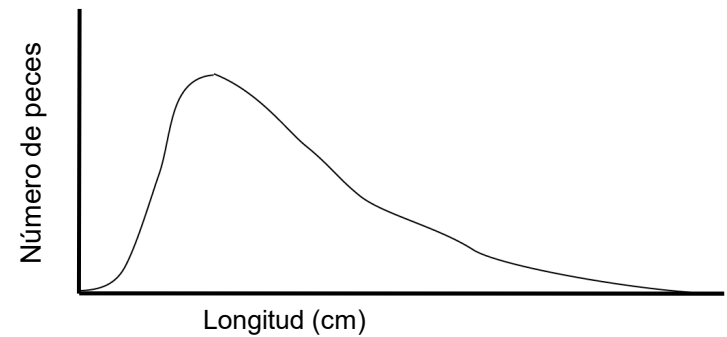


Población target (biológica)

Variable	Indicador
Tamaño	Longitud (cm)



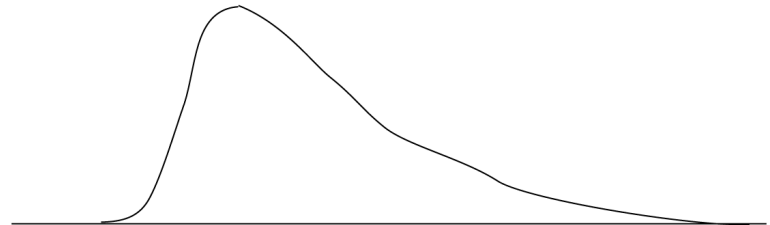
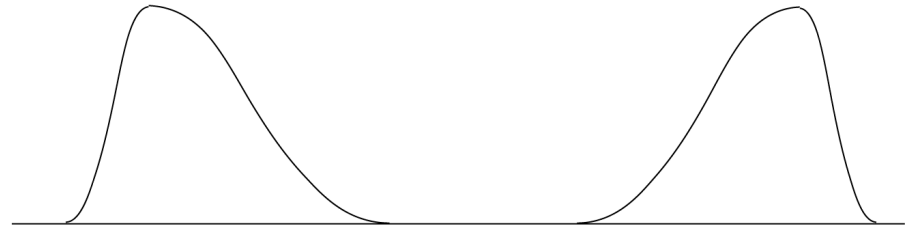
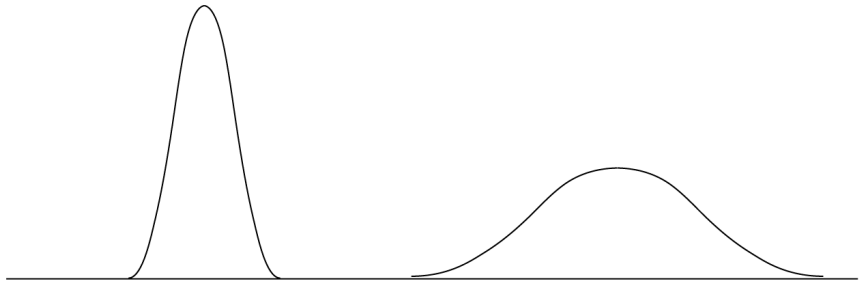
Una variable



Población estadística

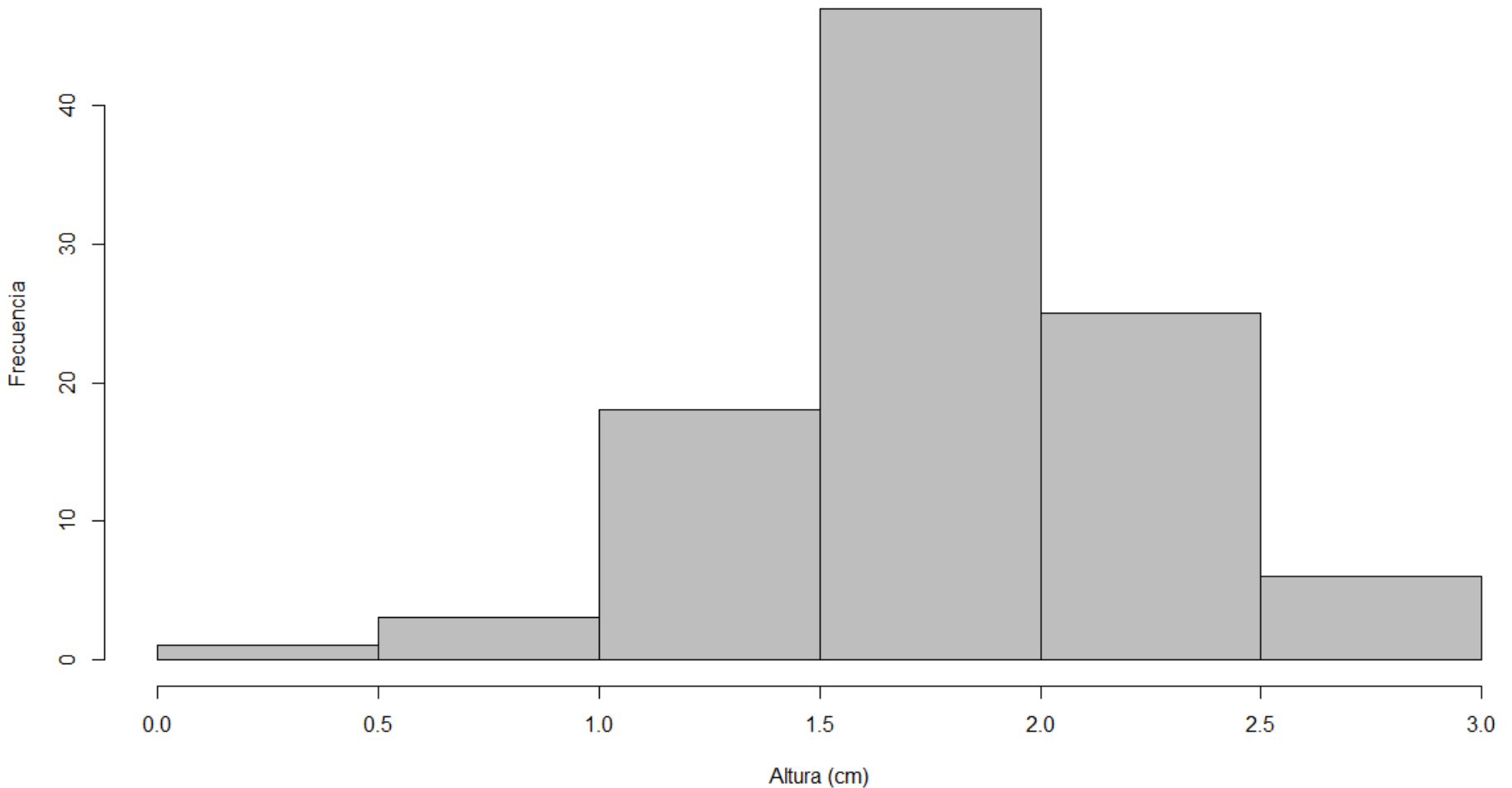
DISTRIBUCIÓN DE FRECUENCIA DE VARIABLES CONTINUAS

1. Mesocúrtica
2. Leptocúrtica
3. Platicúrtica



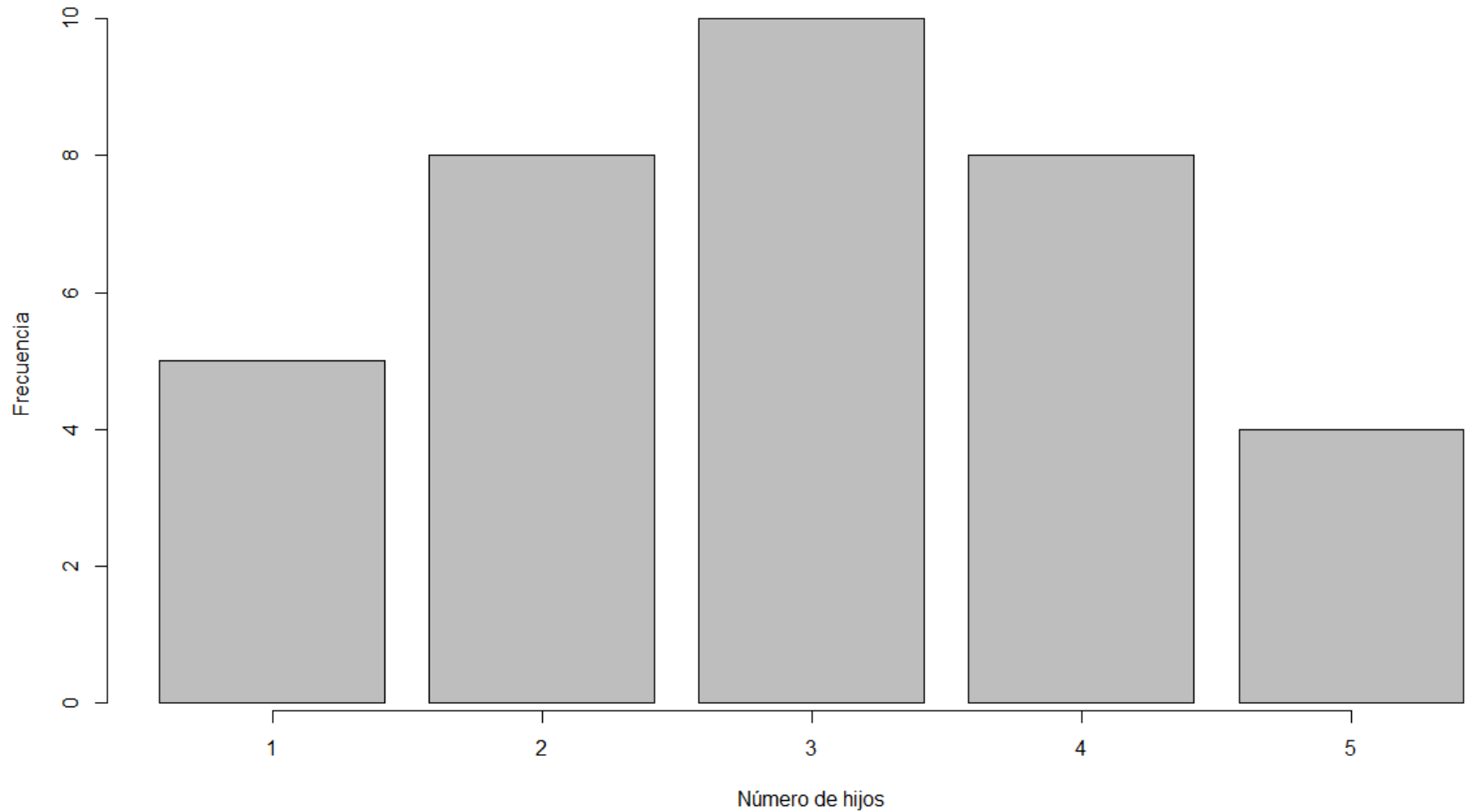
Regularidad... tendencia.... Distribuciones de frecuencia

Variable en escala continua



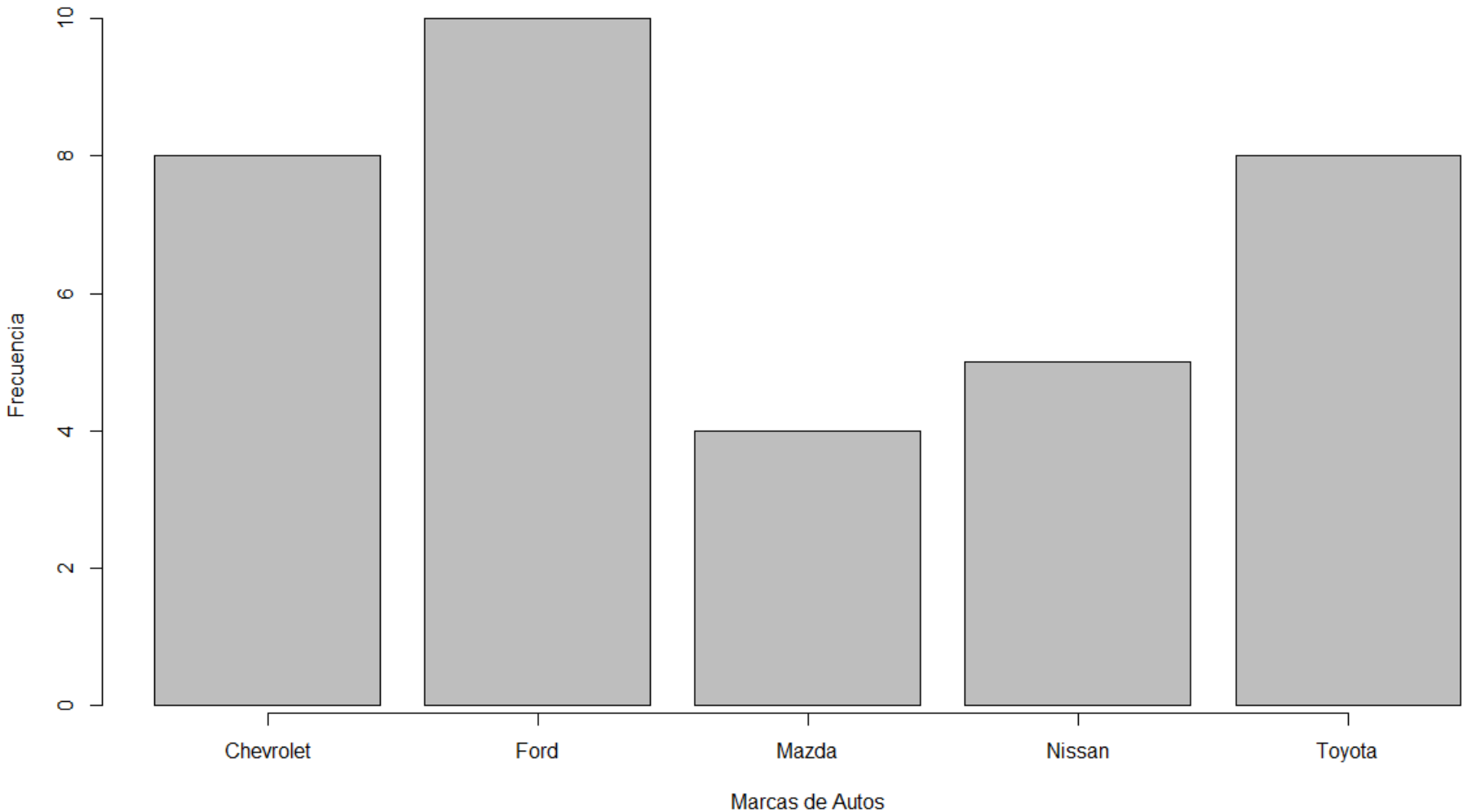
Regularidad... tendencia.... Distribuciones de frecuencia

Variable en escala discreta



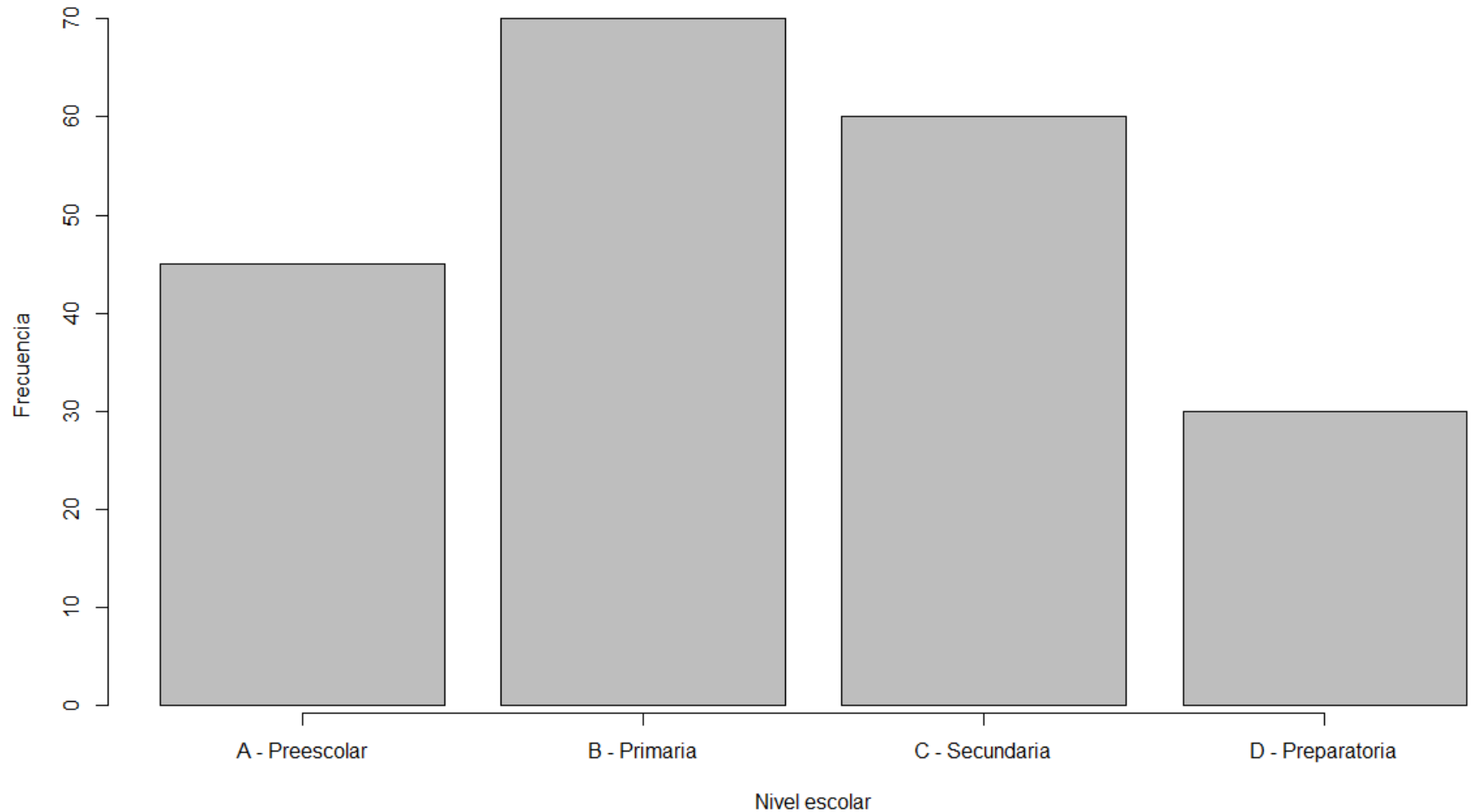
Regularidad... tendencia.... Distribuciones de frecuencia

Variable de atributo en escala nominal



Regularidad... tendencia.... Distribuciones de frecuencia

Variable de rango en escala ordinal



Las distribuciones de frecuencia son una de las posibilidades para representar directamente las propiedades de una población estadística...

- Diagramas de cajas (Boxplot)
- Gráficos de puntos (Scatter plot)
- Gráficos de línea (Line plot)
- Gráficos tipo violín (Violin plot)
- Gráfico de barras (Bar plot)
- Gráfico de promedios (Means plot)



Distribuciones de frecuencias:

Muestran un resumen del número de observaciones que toman valores específicos o intervalos de valores de una variable.

La frecuencia es una cuenta del número de veces que la variable X tomó un dado valor o intervalo de valores.

Tabla 1. Número de góbdos bajo las piedras en la zona de inter-marea rocosa de la costa Norte de Yucatán.

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

¿Cuántos góbdos hay por roca? ¿se agregan?

¿Varía mucho el número de góbdos?

¿La mayoría de las piedras tienen góbdos?

Frecuencias absolutas

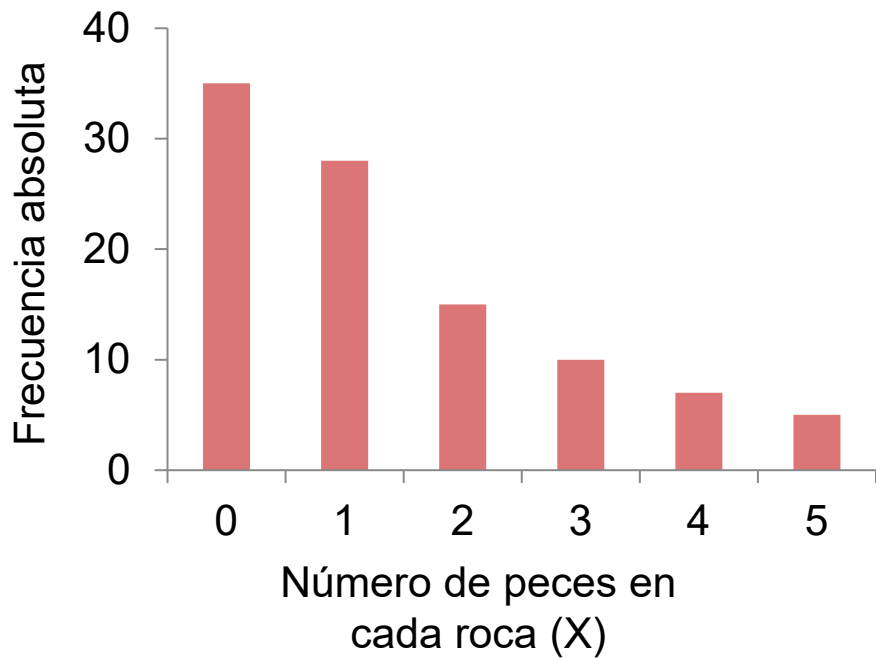
Tabla 2. Frecuencia absoluta y absoluta acumulada de rocas con X número de góbidos en la zona de inter-marea rocosa de la costa Norte de Yucatán.

Número de peces (X)	Frecuencia absoluta	Frecuencia absoluta acumulada
0	35	35
1	28	63
2	15	78
3	10	88
4	7	95
5	5	100
Total	100	

X es una variable discreta

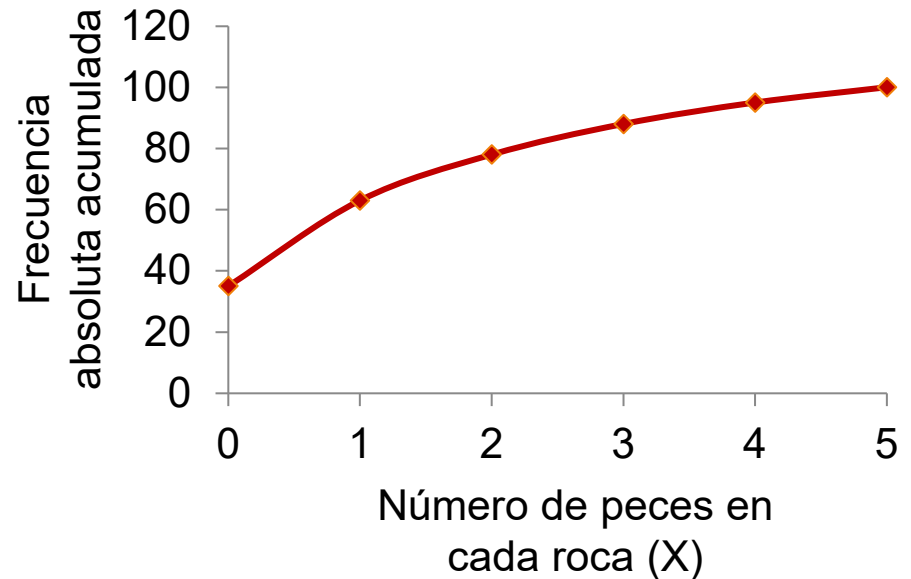
Rápidamente se observa que:

- una tercera parte de las rocas no tuvieron peces;
- que muchos peces se encontraron de uno en uno;
- y que pocas rocas tuvieron agregaciones de mas de 3 peces;



La mayoría de las rocas no tuvieron peces, o bien tuvieron 1; que hubo 4 y 5 peces en aproximadamente el mismo número de rocas.

Casi 80 de las 100 rocas tuvieron 2 peces o menos. Sólo el 10% de las rocas tuvieron 4 y 5 peces.

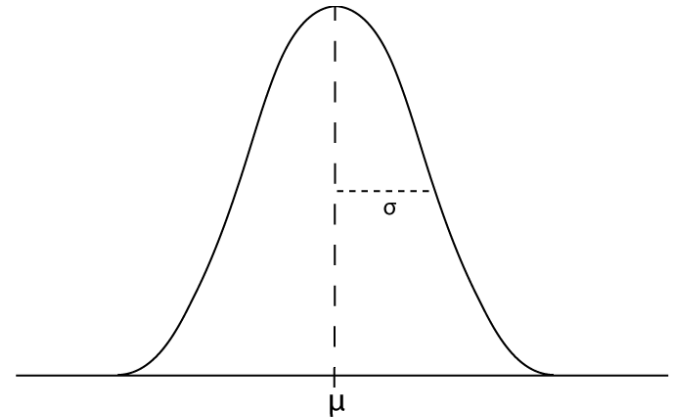


Frecuencias relativas

Tabla 2. Frecuencia absoluta, absoluta acumulada, relativa y relativa acumulada de rocas con X número de góbidos en la zona de inter-marea rocosa de la costa Norte de Yucatán.

Número de peces (X)	Frec. absoluta	Frec. absoluta acumulada	Frec. relativa	Frec. relativa acumulada
0	35	35	0.35	0.35
1	28	63	0.28	0.63
2	15	78	0.15	0.78
3	10	88	0.1	0.88
4	7	95	0.07	0.95
5	5	100	0.05	1
Total	100		1	

DISTRIBUCIÓN DE FRECUENCIA DE VARIABLES CONTINUAS



Parámetro de localización: μ

Desviación estándar: σ

Simetría: Sesgo y Kurtosis

Propiedades de un muestreo

- Preciso
- Certero

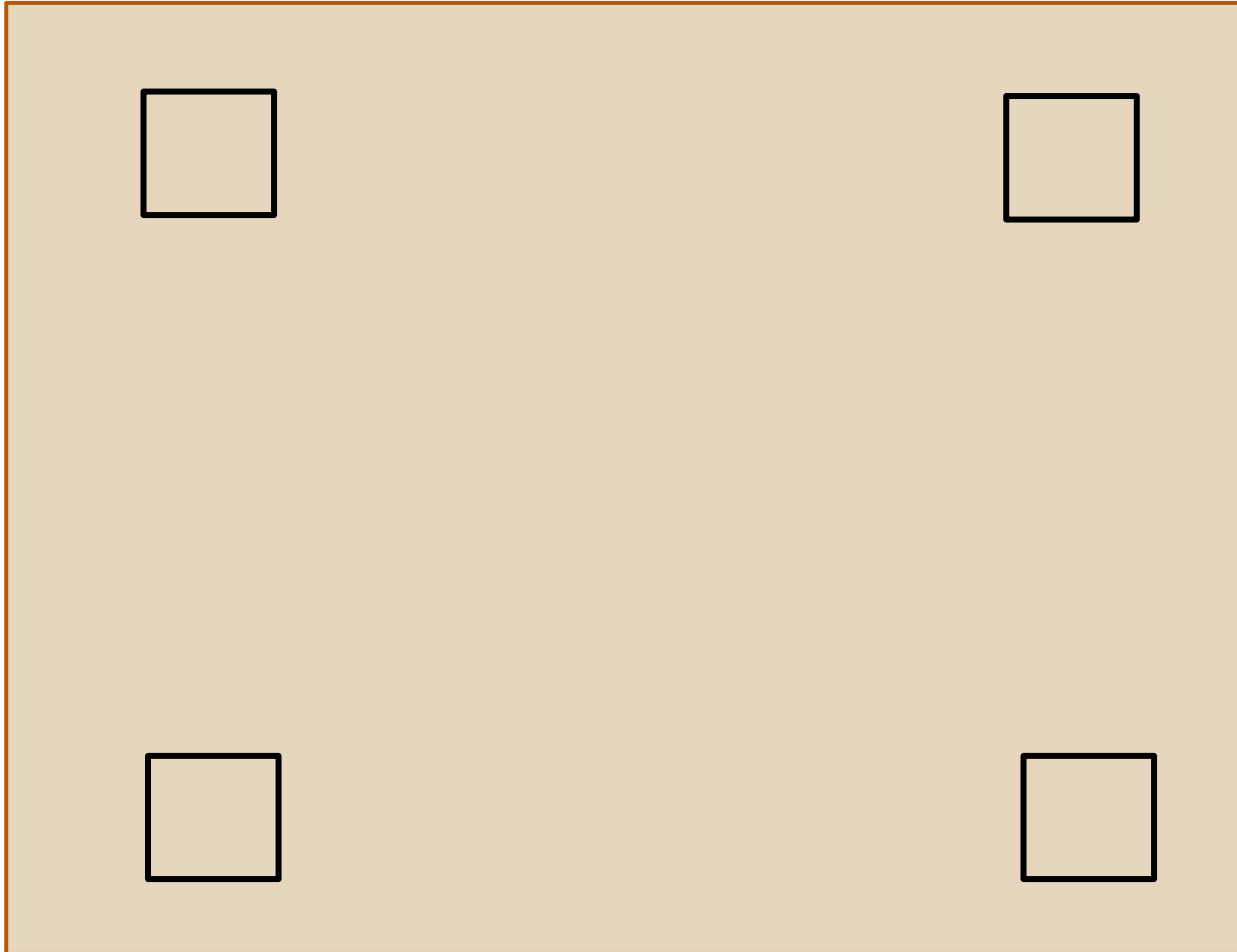


REPRESENTATIVO

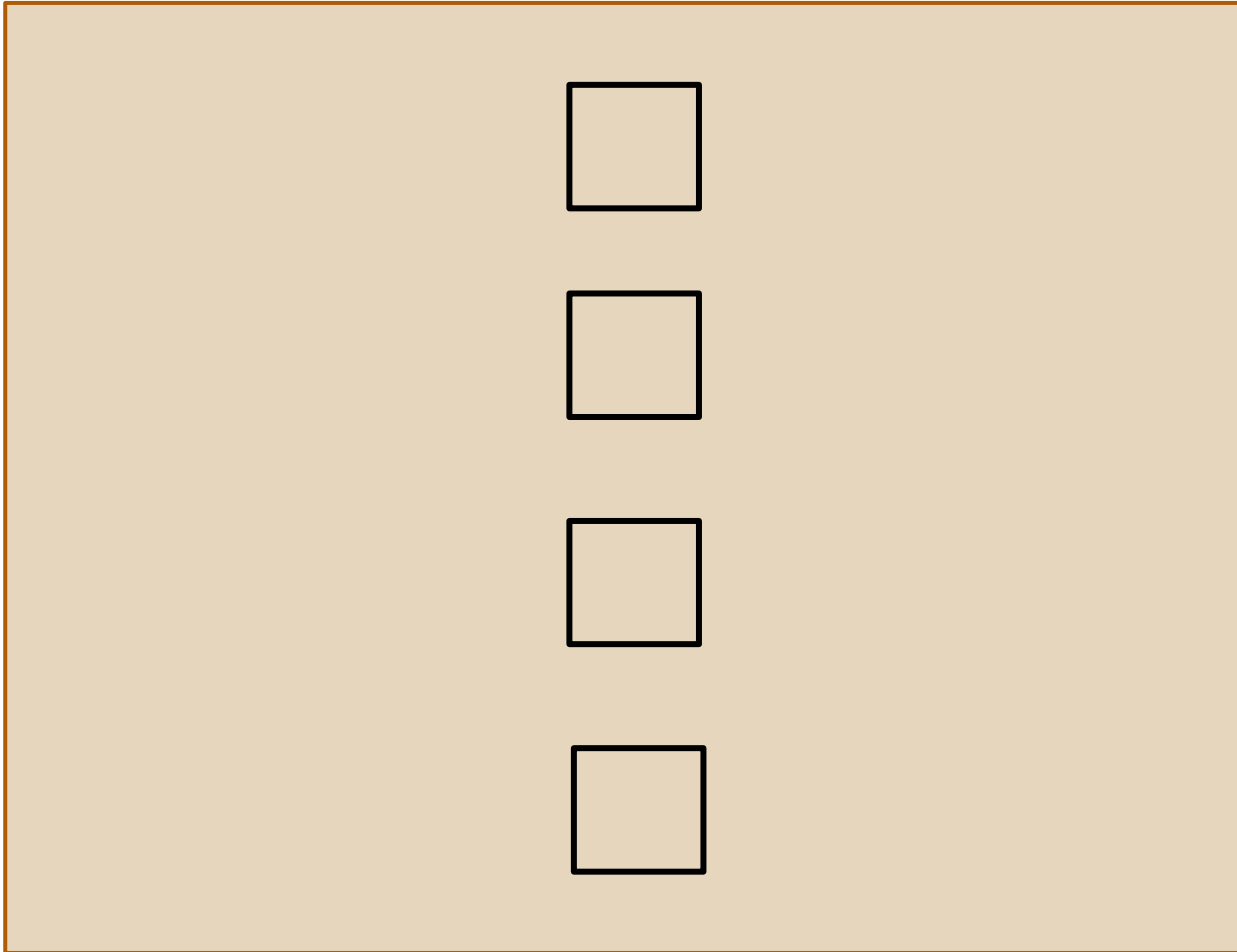
¿cómo obtenemos un muestreo representativo?

- Apropiado tamaño muestral
- Muestreo aleatorio (no garantiza)

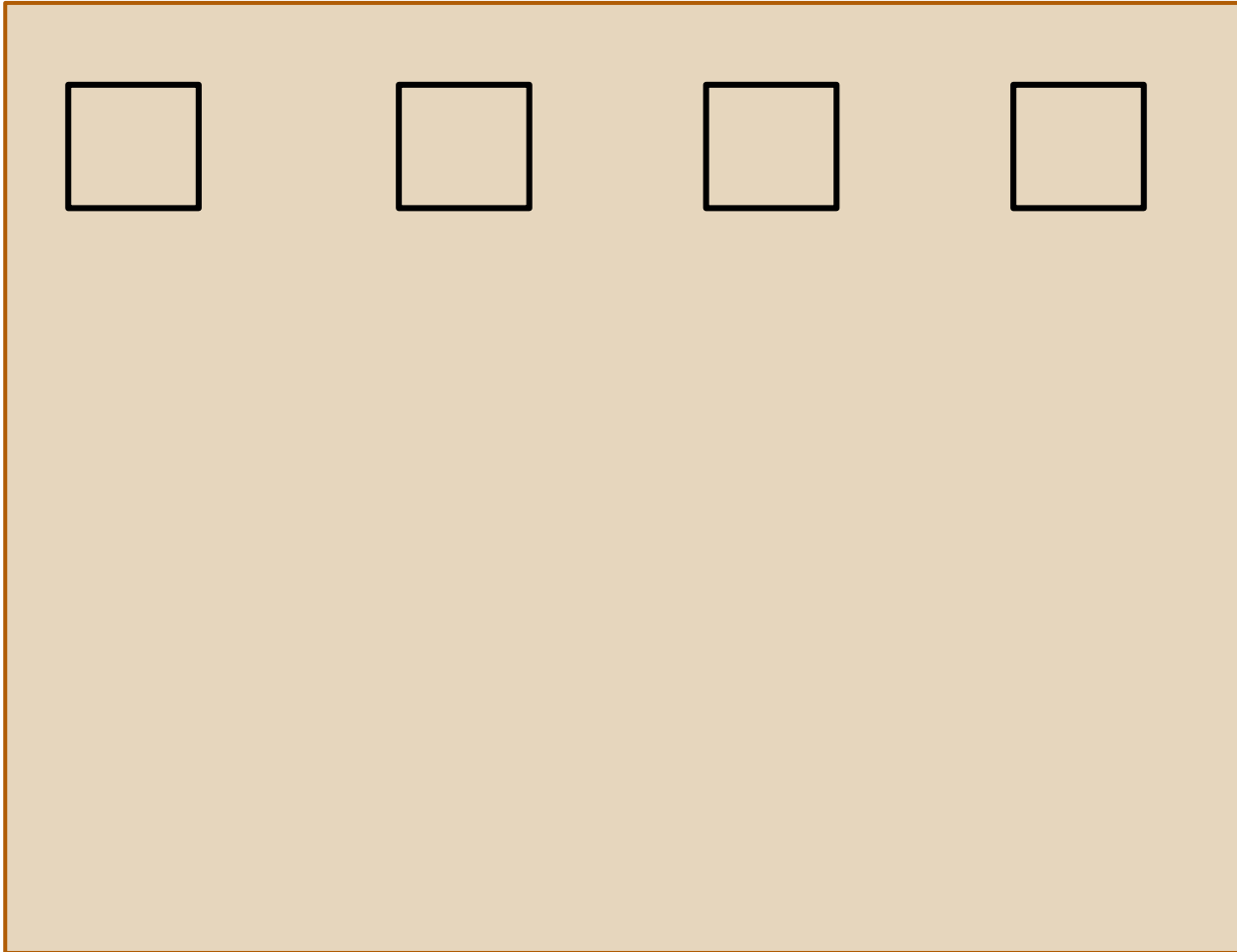
¿Muestreo representativo?



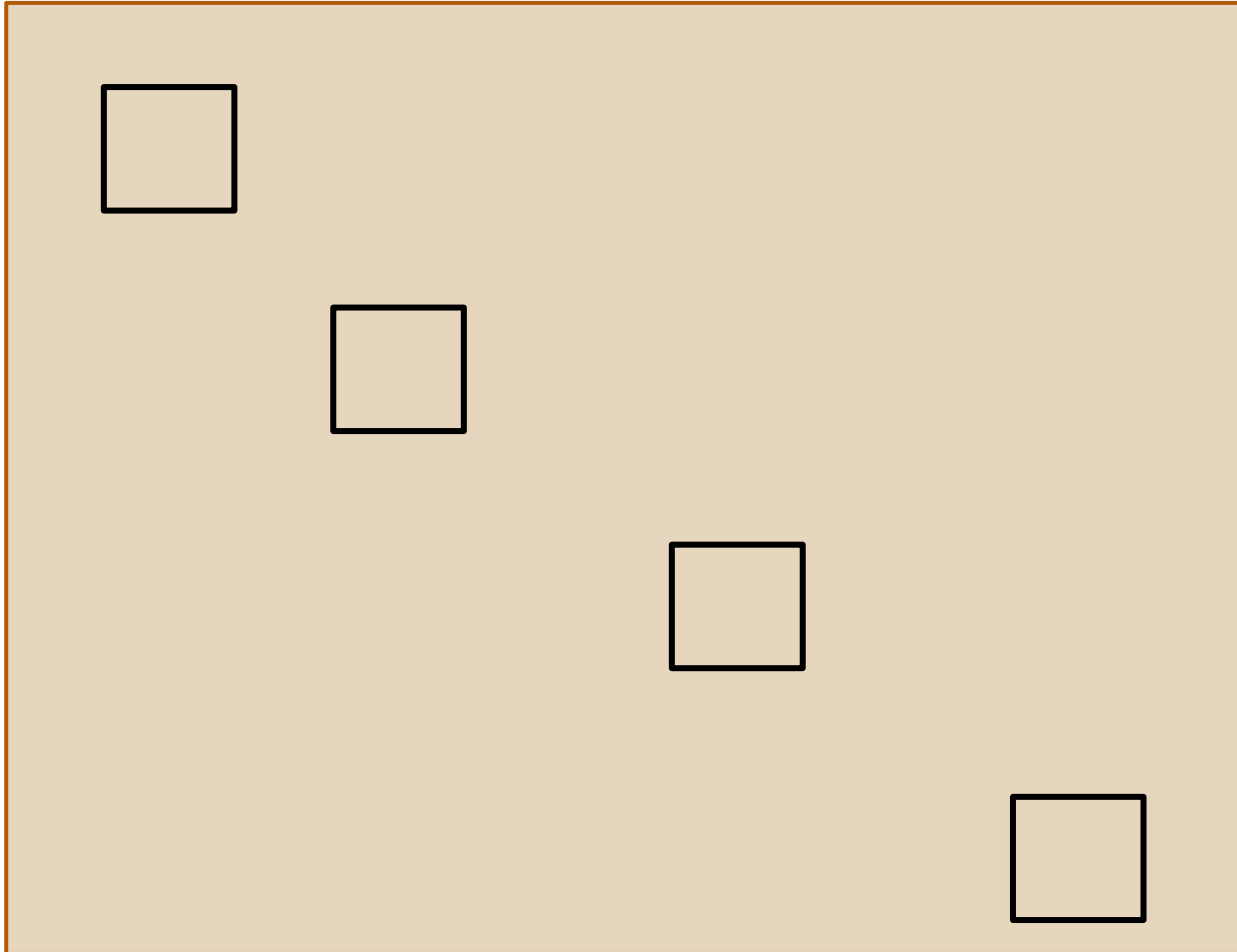
¿Muestreo representativo?



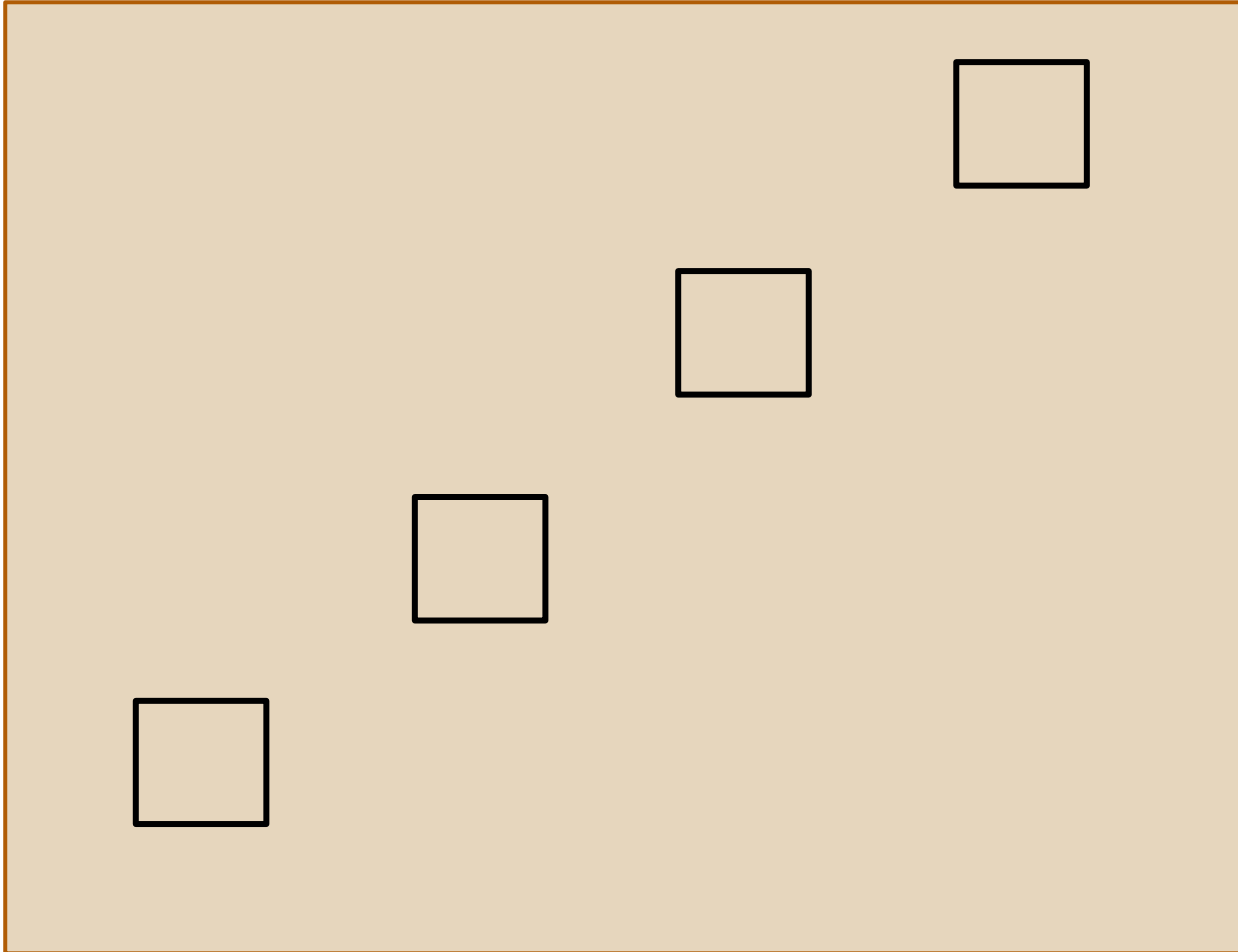
¿Muestreo representativo?



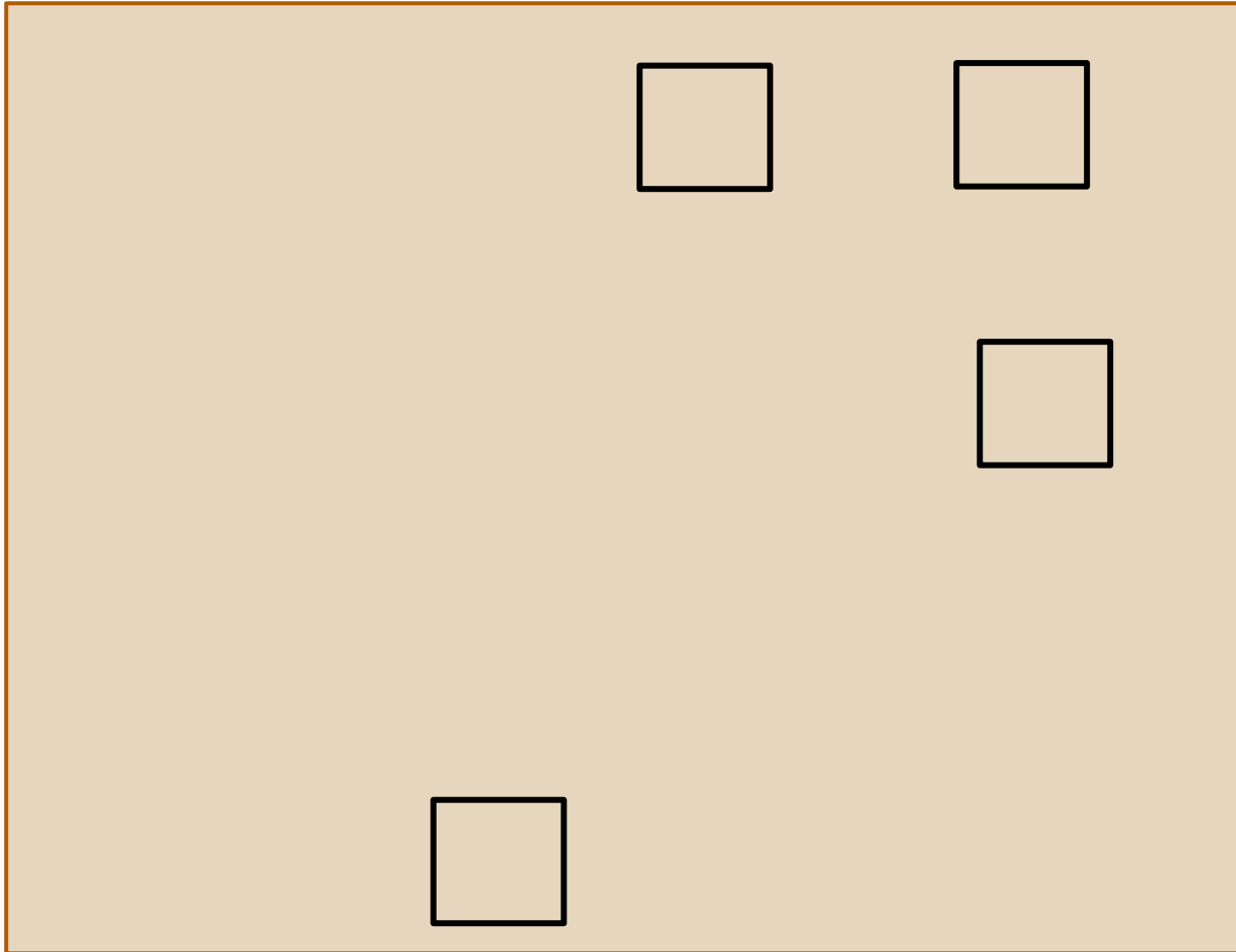
¿Muestreo representativo?



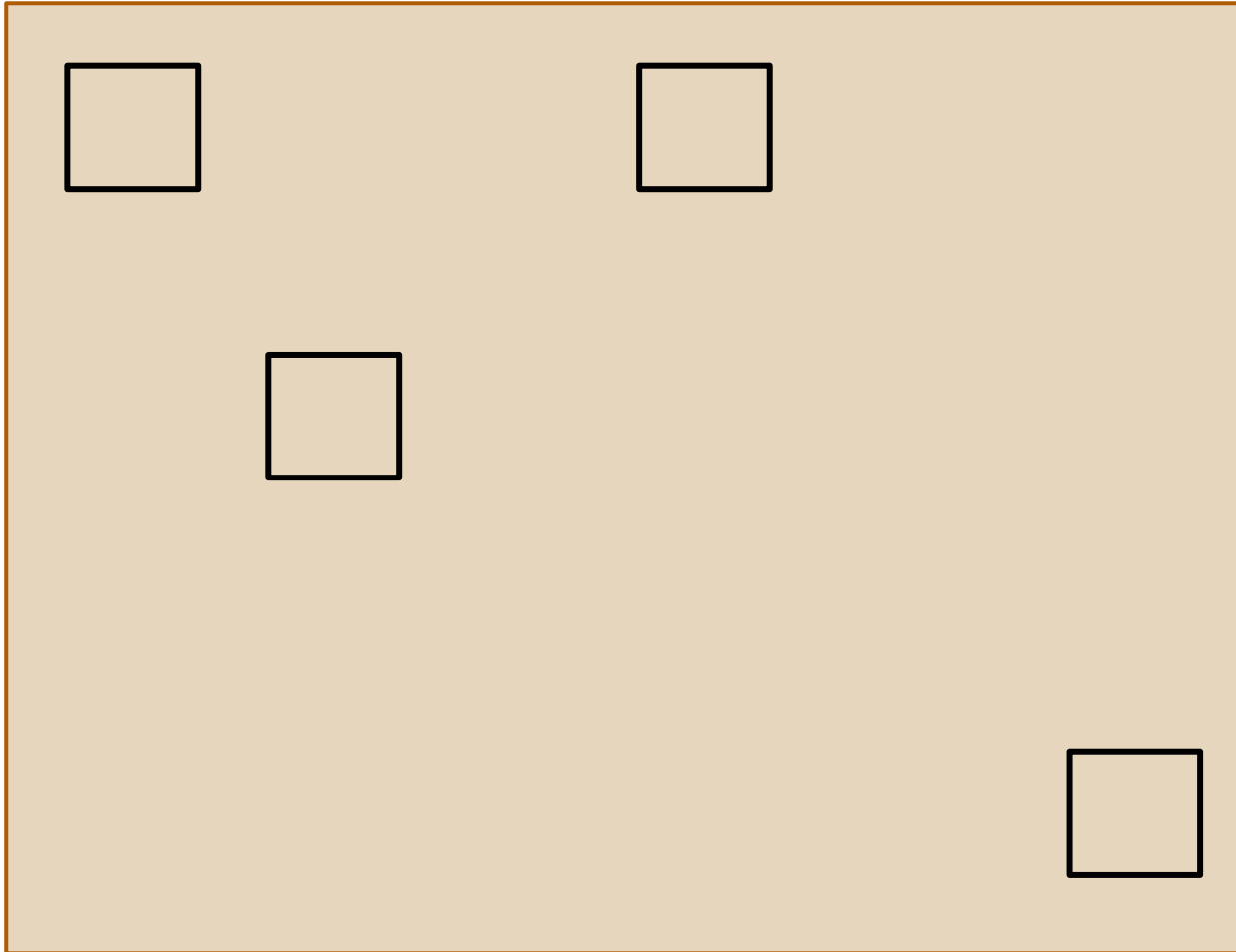
¿Muestreo representativo?



¿Muestreo representativo?

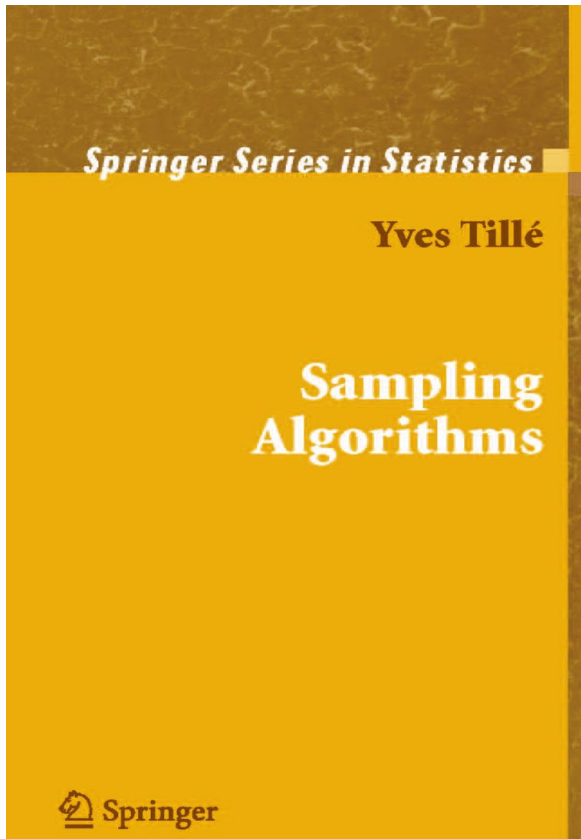


¿Muestreo representativo?

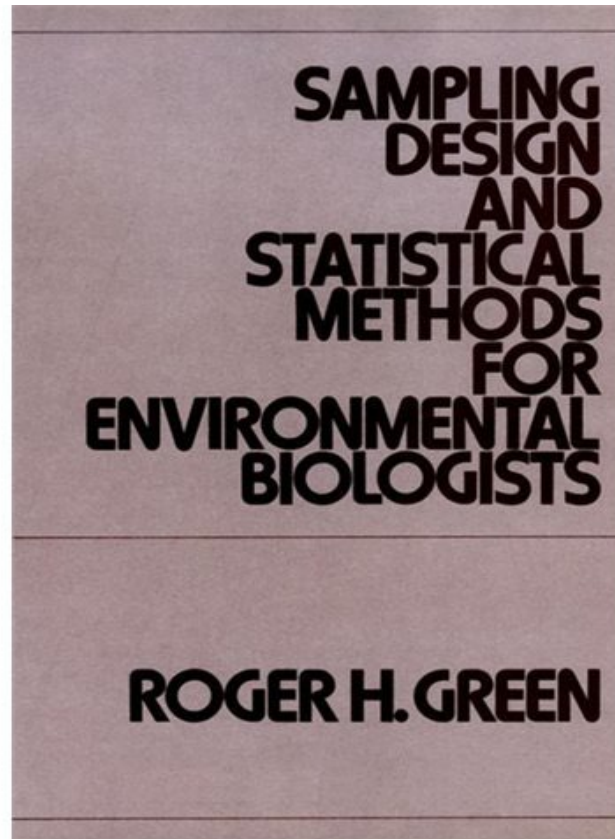


Tipos de Muestreo:

1. Muestreo aleatorio simple
2. Muestreo aleatorio sistemático
3. Muestreo ordenado sistemático
4. Muestreo aleatorio estratificado
5. Muestreo aleatorio de doble etapa



46 procedimientos de muestreo
Paquete “sampling” en R



Clásico en Ciencias Ambientales (propuso BACI)

TENDENCIA CENTRAL

Resumen el nivel o punto central de una distribución de frecuencias.

Media (μ)

Mediana

Quantiles

Moda

DISPERSIÓN Y FORMA

Indican cuan dispersos o juntos están los datos alrededor de un valor central. Medida de la variabilidad.

Rango

Desviación estándar (σ)

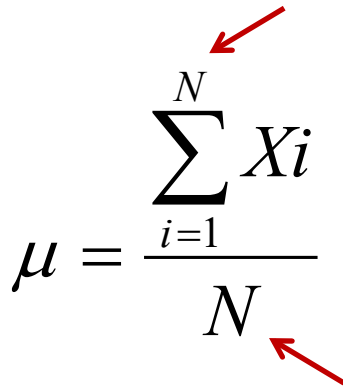
Varianza (σ^2)

Coeficiente de variación

Curtosis y Simetría

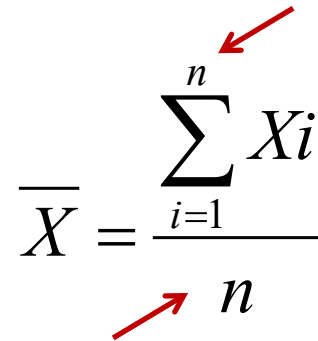
Media aritmética

La media aritmética es el parámetro de tendencia central mas conocido, y comúnmente se le llama media o promedio. Es el valor que resume el valor de la variable X.

$$\mu = \frac{\sum_{i=1}^N Xi}{N}$$


POBLACIÓN

El mejor estimador
(no sesgado) de μ es:

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


MUESTRA

Mediana

La mediana se define como el valor central de una serie de valores ordenados. Es decir hay el mismo número de observaciones menores que la mediana que mayores que la mediana.

La median se obtiene ordenando los valores de X_i de menor a mayor y calculando:

$$M = X_{(N + 1)/2}$$



POBLACIÓN

El mejor estimador
de la mediana es:

$$M = X_{(n + 1)/2}$$



MUESTRA



En una población donde N es un número entero impar, $(N + 1) / 2$ será un número entero. En una población donde N es un número par, $(N + 1) / 2$ y M será el valor medio entre los dos valores de M correspondientes.

Cuantiles

Los cuantiles son valores de una serie de datos ordenados que la separan en diversas fracciones. De la misma manera que la mediana la divide a la mitad, los cuantiles la dividen en, por ejemplo, el 30 % mas bajo de los datos.

Cuando se trata de dividir los datos en cuatro partes iguales de 25%, se habla de cuartiles. El primer cuartil (Q_1) es el 25% mas bajo de los datos; el segundo cuartil (Q_2) es la mediana; el tercer cuartil (Q_3) es el primer 25% mas el segundo 25% mas el tercer 25%, o sea el 75% mas bajo de los datos.

$$Q_1 = X_{(N+1)/4}$$

$$Q_3 = X_{(N+1) - Q_1}$$

POBLACIÓN

El mejor estimador
de un cuartil es:

$$Q_1 = X_{(n+1)/4}$$

$$Q_3 = X_{(n+1) - Q_1}$$

MUESTRA

Moda

La moda se define como el valor de la variable más frecuente. Existen distribuciones de frecuencias con una moda, dos modas (bi-modal), o mas modas (multimodal), o bien, sin moda.

$$m = X_{f(x) \text{ maxima}}$$

POBLACIÓN

El mejor estimador
de la moda es:


$$m = X_{f(x) \text{ maxima}}$$

MUESTRA

- ★ Cuando se tienen dos modas separadas por una serie de valores de X_i , entonces se puede hablar de una moda mayor y una menor, refiriéndose como mayor y menor al valor de la frecuencia y no al valor de X_i . Cuando se tienen dos modas iguales juntas, la moda es el valor intermedio entre los dos valores de X_i consecutivos o contiguos.


Rango

El rango es la diferencia entre el valor máximo y el mínimo de la variable X_i . Si los datos son ordenados de menor a mayor (como con la mediana), el rango se define como:


$$\text{Rango} = X_N - X_1$$


POBLACIÓN

El mejor estimador
de rango es:


$$\text{Rango} = X_n - X_1$$


MUESTRA

$$\text{RIQ} = X_{Q3} - X_{Q1}$$


POBLACIÓN

El mejor estimador
de rango
intercuartil es:

$$\text{RIQ} = X_{Q3} - X_{Q1}$$


MUESTRA

Varianza

Es una medida de dispersión que provee información sobre la magnitud de las distancias promedio de todos los valores X_i de la media; es decir, nos dice cuan dispersos están los datos alrededor de un parámetro de tendencia central.

$$\sigma^2 = \frac{\sum (X_i - \mu)^2}{N}$$

POBLACIÓN

El mejor estimador
de σ^2 es:

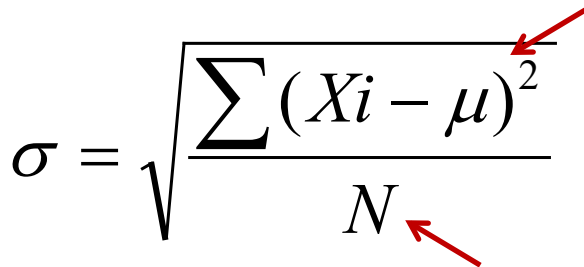
$$S^2 = \frac{\sum (X_i - \bar{X})^2}{n - 1}$$

MUESTRA

- ★ El denominador en la varianza poblacional es N mientras que para la muestral es $n-1$. Esto es porque si se utiliza N para calcular el estadístico de la varianza se obtienen valores sesgados de la varianza poblacional. Por ello se divide entre los grados de libertad ($gl = n-1$).

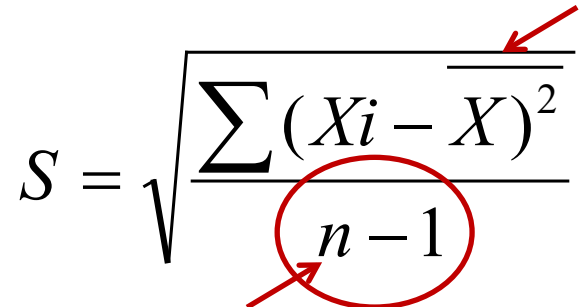
Desviación estándar

Las unidades de todos los parámetros vistos hasta ahora son las mismas unidades que las de la variable, excepto las de la varianza que son las unidades de X_i al cuadrado. Para poder tener un parámetro de dispersión en las mismas unidades que la variable, se utiliza la desviación estándar.

$$\sigma = \sqrt{\frac{\sum (X_i - \mu)^2}{N}}$$


POBLACIÓN

El mejor estimador
de σ es:

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


MUESTRA

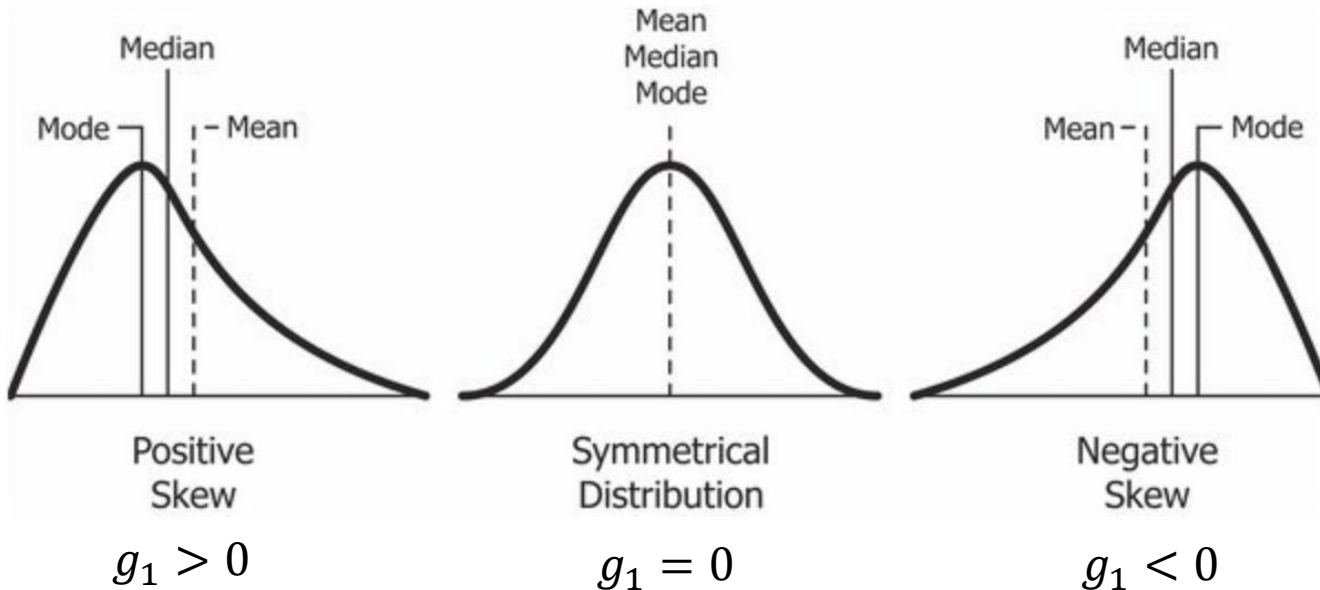
- ★ El denominador en la desviación estándar poblacional es N mientras que para la muestral es $n-1$. Esto es porque si se utiliza N para calcular el estadístico de la varianza se obtienen valores sesgados de la varianza poblacional. Por ello se divide entre los grados de libertad ($gl = n-1$).

Simetría

Permite identificar si los datos se distribuyen de forma uniforme alrededor del promedio.

$$\tilde{\mu}_3 = \frac{\mu_3}{\sigma^3}$$

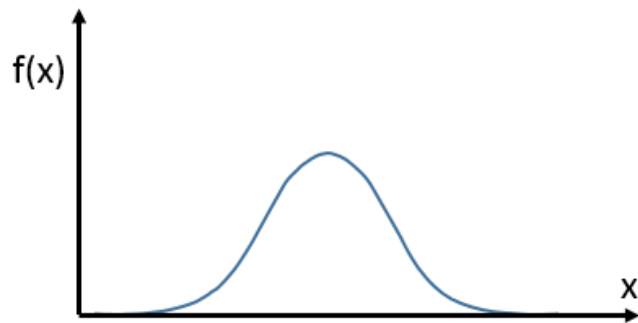
$$g_1 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{3/2}}$$



Curtosis (kurtosis en inglés)

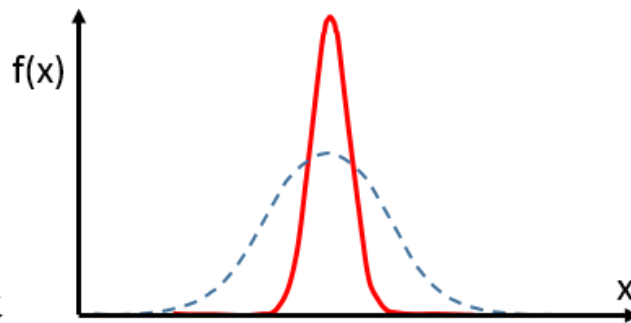
Esta medida determina la propensión a generar valores extremos respecto al promedio según el grado de concentración que presentan los valores en la región central de la distribución.

$$\tilde{\mu}_4 = \frac{\mu_4}{\sigma^4}$$
$$g_2 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^2}$$



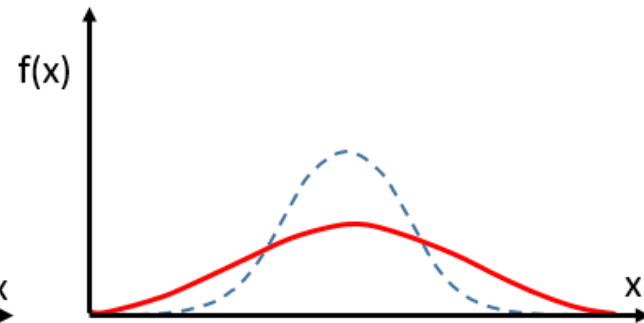
mesocúrtica

$$g_2 = 3$$



leptocúrtica

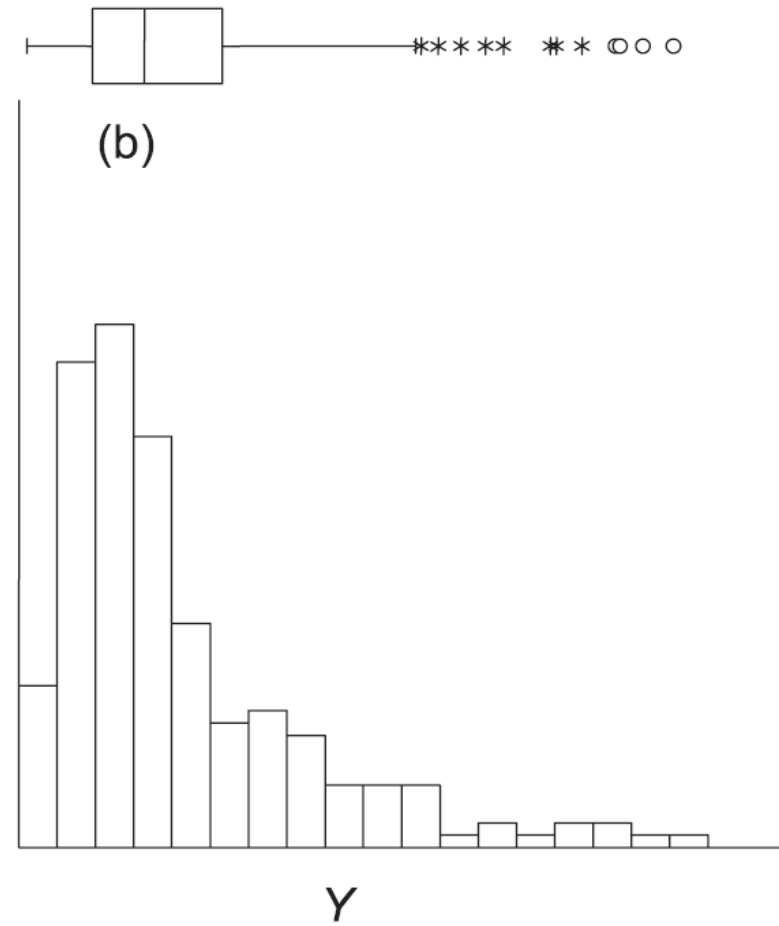
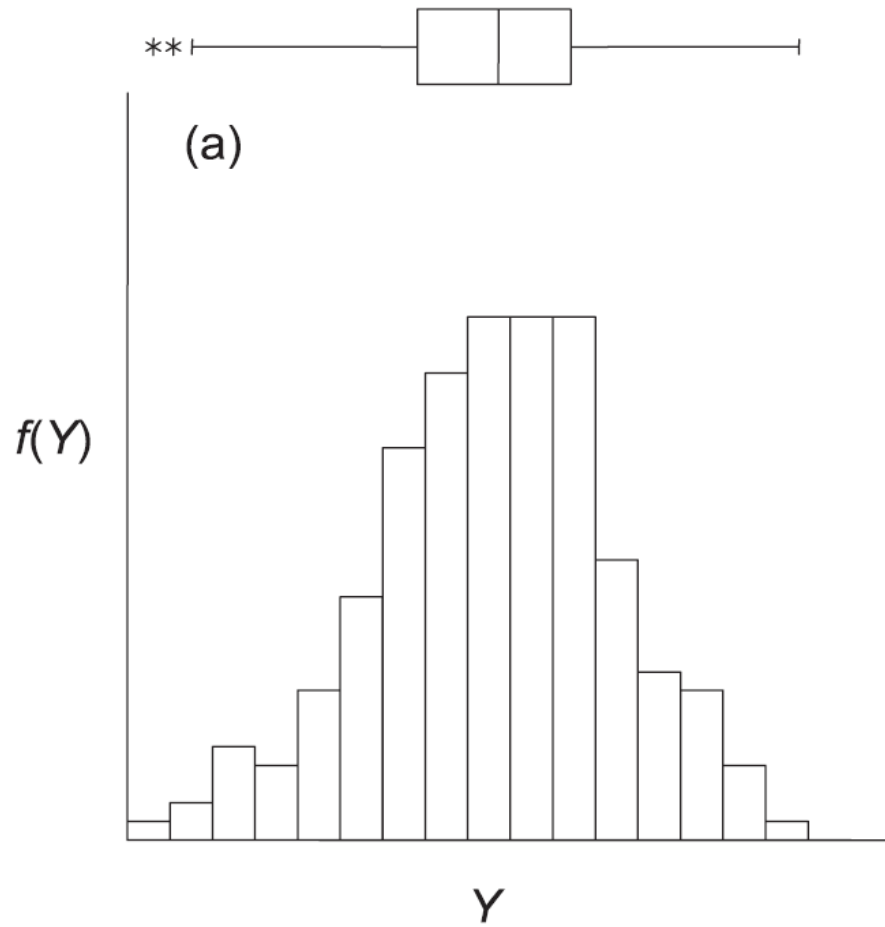
$$g_2 > 3$$

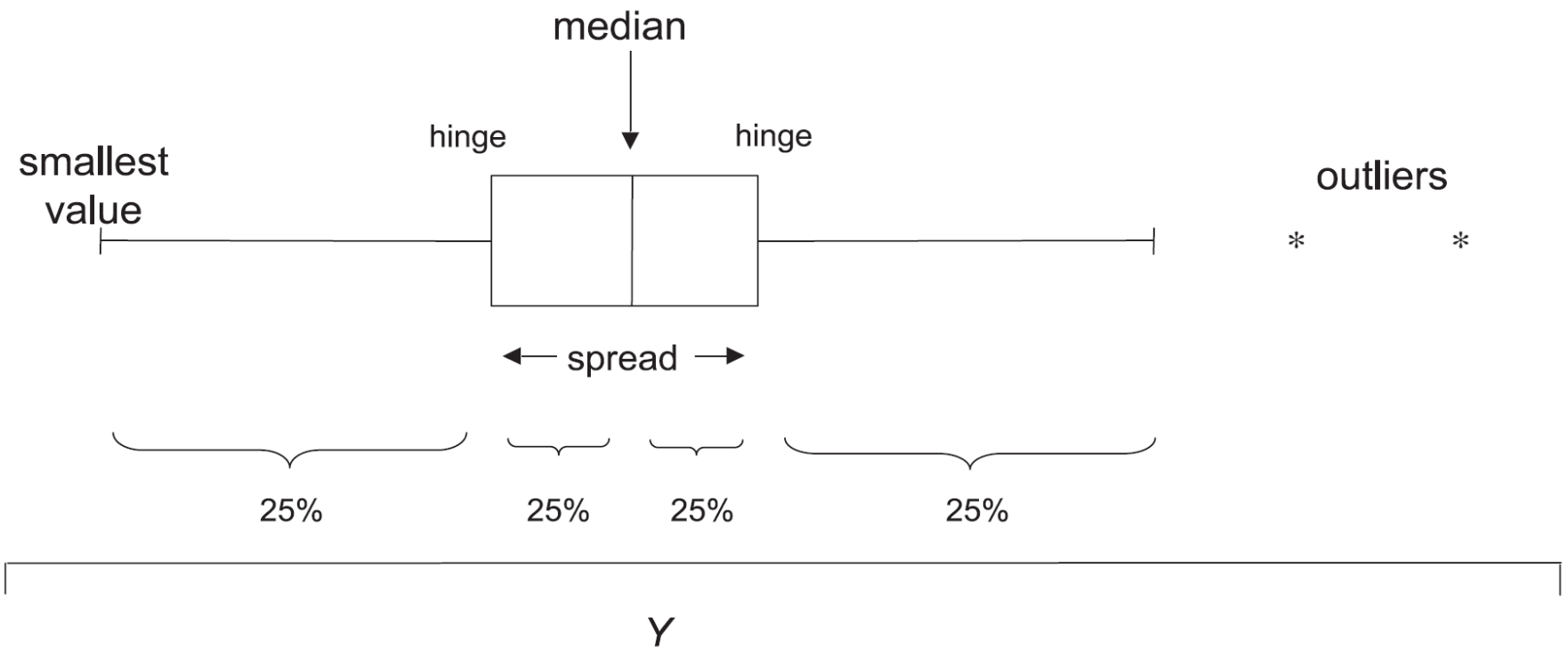


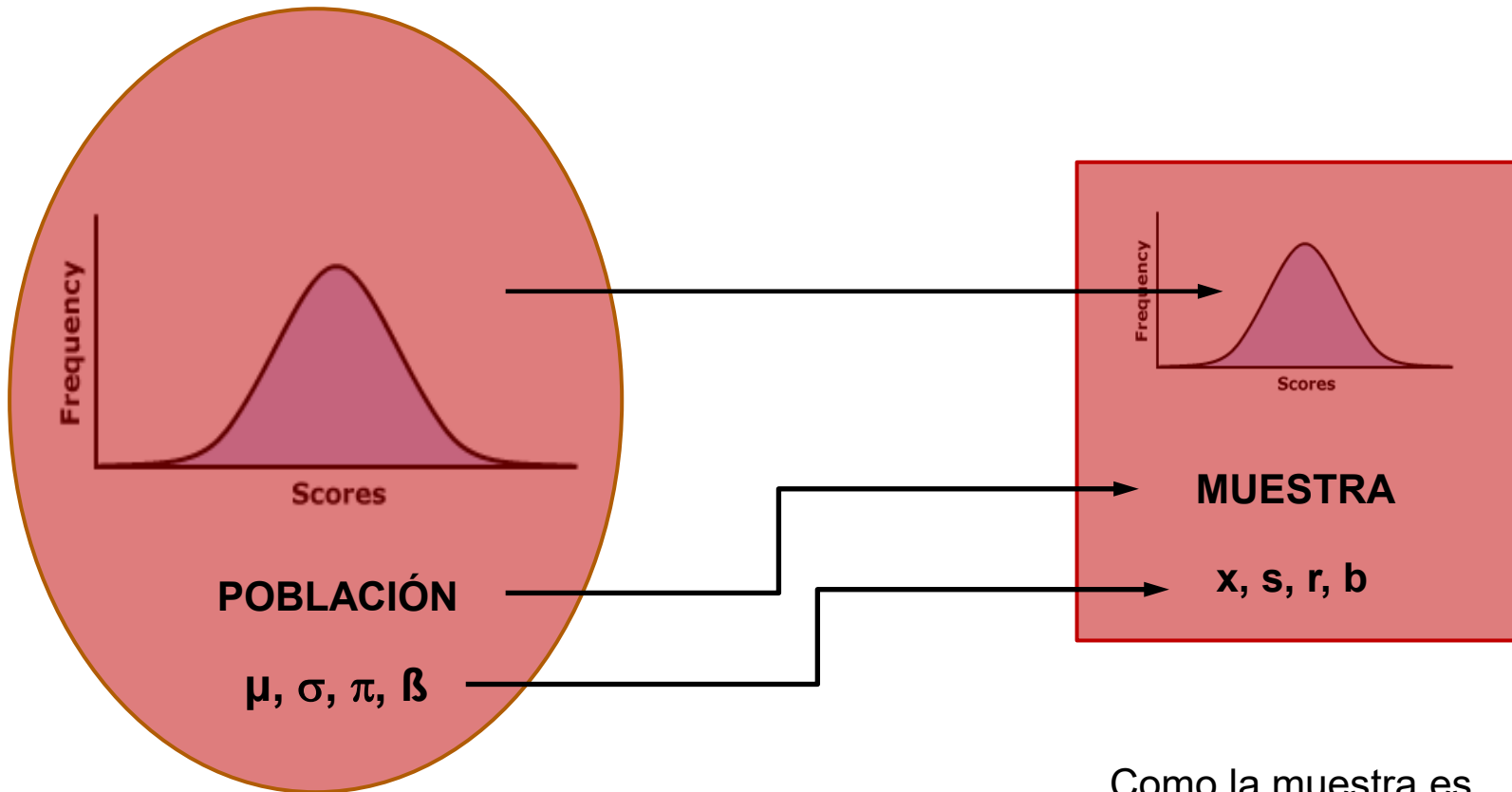
platicúrtica

$$g_2 < 3$$

Exploración gráfica de datos: Diagrama de cajas (Boxplot)







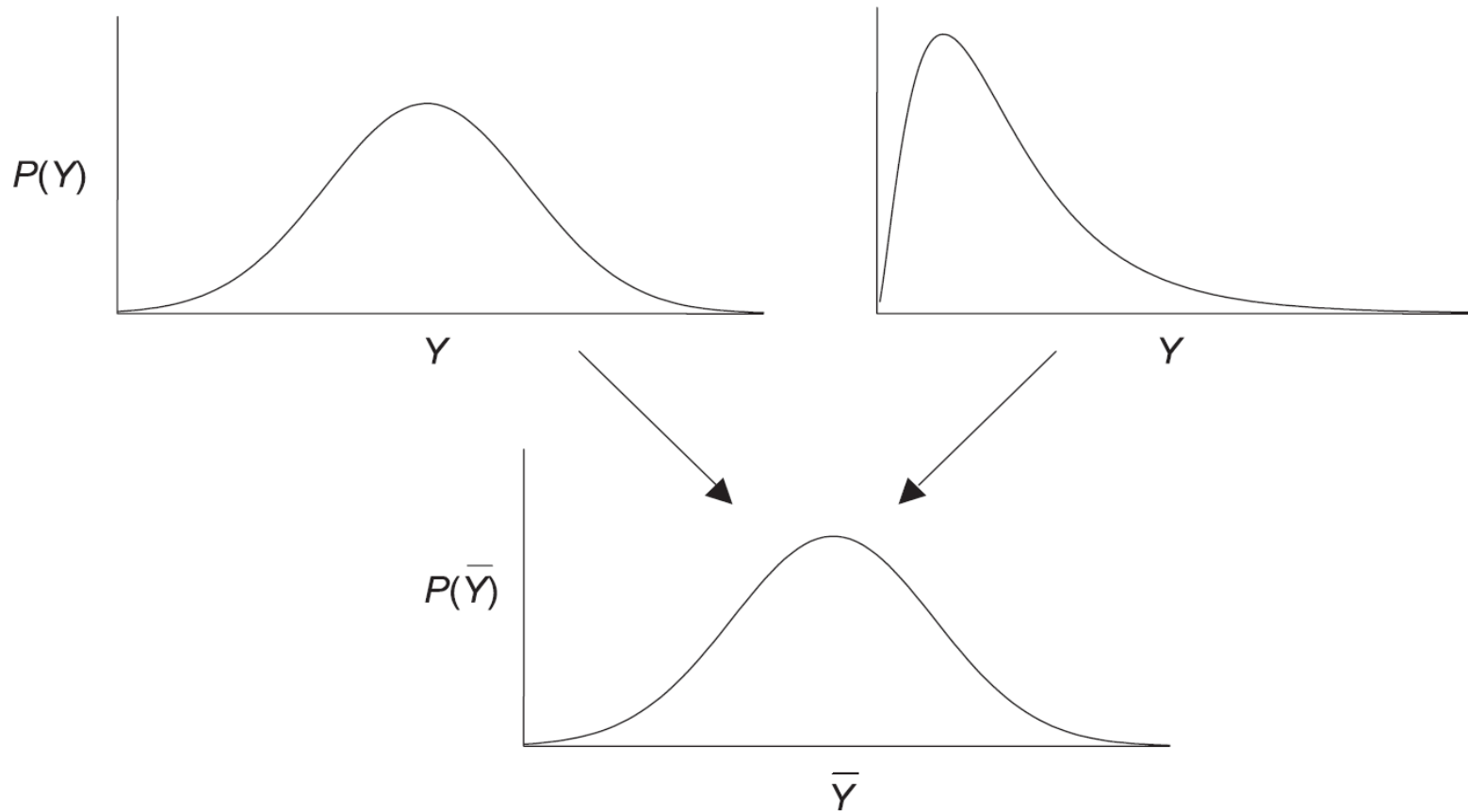
La estadística se ocupa de estimar los parámetros de una distribución de frecuencias de una variable en una población.

MUESTRA



Como la muestra es representativa de la población, la variable en la muestra tendrá una distribución de frecuencias similar a la de la población.

Principio del Teorema Central del límite



$$s_{\bar{y}} = \frac{s}{\sqrt{n}}$$

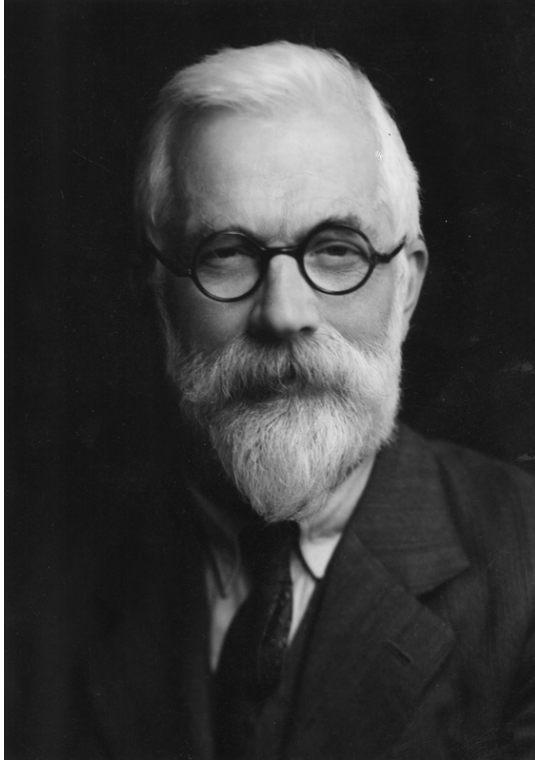
Desviación Estándar de promedios o **Error Estándar** del promedio

La necesidad de la Prueba estadística

Los muestreos están sujetos a error muestral y a la variabilidad natural de las cosas biológicas

Debemos estimar la probabilidad asociada con la relación entre nuestras muestras y la población de la que proviene.

Estamos lidiando con respuestas probabilísticas, no con conclusiones absolutas



Ronald A. Fisher
1890-1962

«llamar a un estadístico después de que el experimento (o muestreo) está listo, no es más que invitarlo a realizar una evaluación post mortem: él estará en capacidad de decir porqué el experimento murió»

R.A. Fisher
Indian Statistical Congress, Sankhya, 1938