



***Imaginación Estadística y descripción de datos***

Programa : Estadística aplicada I





Universidad Autónoma de Estado de México  
Facultad de Economía  
Maestría en Sustentabilidad, Desarrollo Regional y Metropolitano

Material didáctico	Visual
Título	Imaginación Estadística y descripción de datos
Espacio Académico	Facultad de Economía-Arquitectura
Plan de estudios	Maestría en Sustentabilidad , Desarrollo Regional y Metropolitano
Unidad de Aprendizaje	Estadística Aplicada I
Núcleo de formación	
Modalidad	Presencial
Tipo	Obligatoria





**Este material se compone de 68 diapositivas para apoyar el desarrollo de la unidad de aprendizaje “ Estadística Aplicada I” de la Maestría en Sustentabilidad y Desarrollo Regional cuyos contenidos temáticos del modulo1: “Nivel Conceptual y Teórico de la Estadística Descriptiva e Inferencial” son los siguientes:**

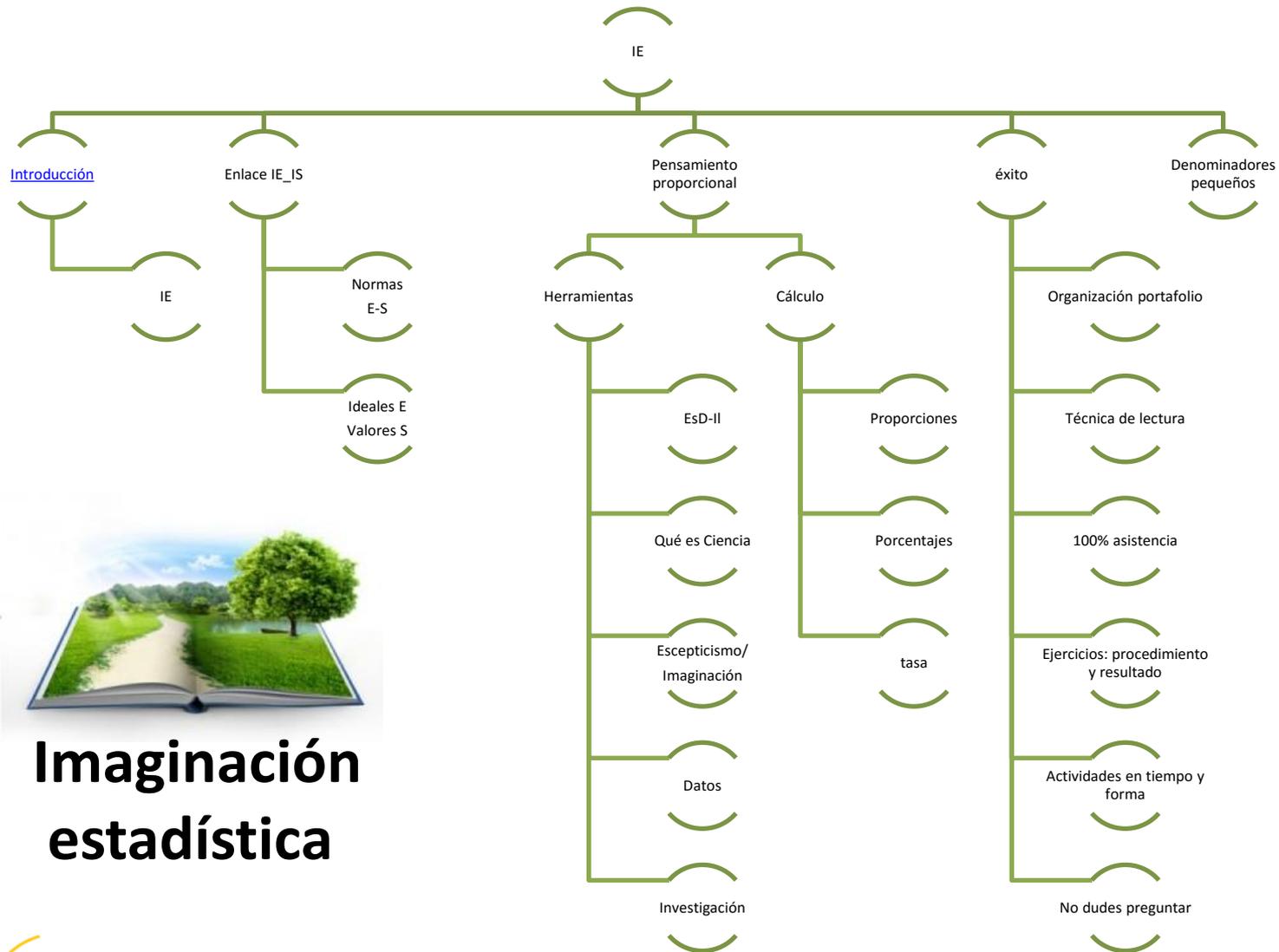
- 1. Imaginación Estadística**
- 2. Descripción de datos: Grafica y Numérica**

**Cada diapositiva se encuentra ordenada por temas de conformidad al programa y calendario del curso : el primer tema comprende las diapositivas de la 4 a la 25; el segundo de la 26 a la 39 y el tercero de la 40 a la 67. La diapositiva 68 corresponde a la bibliografía.**





Universidad Autónoma de Estado de México  
Facultad de Economía  
Maestría en Sustentabilidad, Desarrollo Regional y Metropolitano



# Imaginación estadística





**El campo de la estadística es un conjunto de procedimientos para reunir, medir clasificar, computar , analizar y resumir información numérica adquirida sistemáticamente.**

**“La estadística implica aprender una nueva manera de ver las cosas”**



**“El análisis estadístico es una parte vital del método científico”**

**Mantener un sentido claro de equilibrio y proporción con respecto a la realidad**

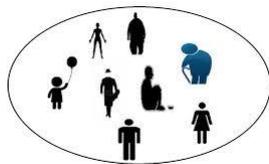




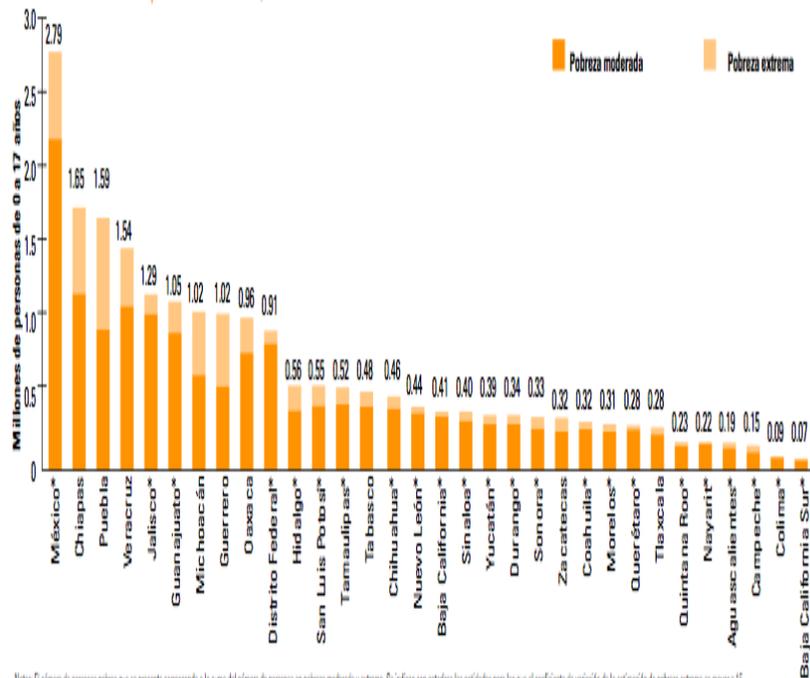
# Universidad Autónoma de Estado de México

## Facultad de Economía

### Maestría en Sustentabilidad, Desarrollo Regional y Metropolitano



GRÁFICA 10. Número de personas de 0 a 17 años en situación de pobreza y pobreza extrema, según entidad federativa de residencia. Millones de personas. México, 2012



Notas: El número de personas pobres que se presenta corresponde a la suma del número de personas en pobreza moderada y extrema. Se indican con asterisco las entidades para las que el coeficiente de variación de la estimación de pobreza extrema es mayor a 15. Fuente: Estimaciones propias a partir de la información del MCS-ENIGH 2012.

*Imaginación sociológica*



## Conocimiento de la relación del individuo con la sociedad y con la historia.\*



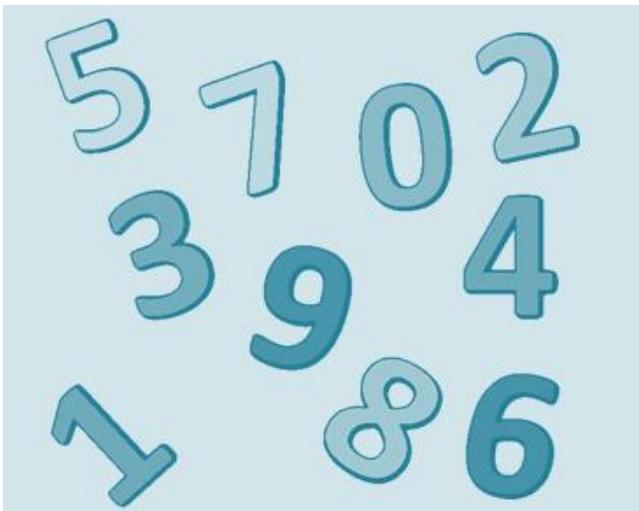
- El comportamiento individual se rige en función de estructuras sociales más grandes;
- Las acciones individuales deben apegarse a las reglas de la sociedad y no a iniciativa personal;
- Las reglas se definen dentro de un contexto cultural.
- Ver el árbol y el bosque.

*\*Charles Wright Mills (28 de agosto de 1916, Waco (Texas) – 20 de marzo de 1962, West Nyack, (Nueva York) sociólogo estadounidense.*





Σ



La Imaginación

*Estadística*





**Apreciación de qué tan usual es un evento, Circunstancia o Comportamiento, en Relación con un Conjunto mayor de Eventos similares.**

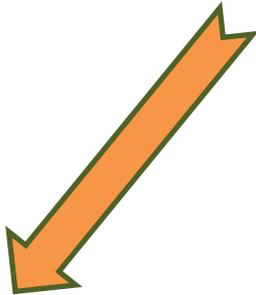


**Y una apreciación de Las causas y Consecuencias del evento, Circunstancia o Comportamiento**

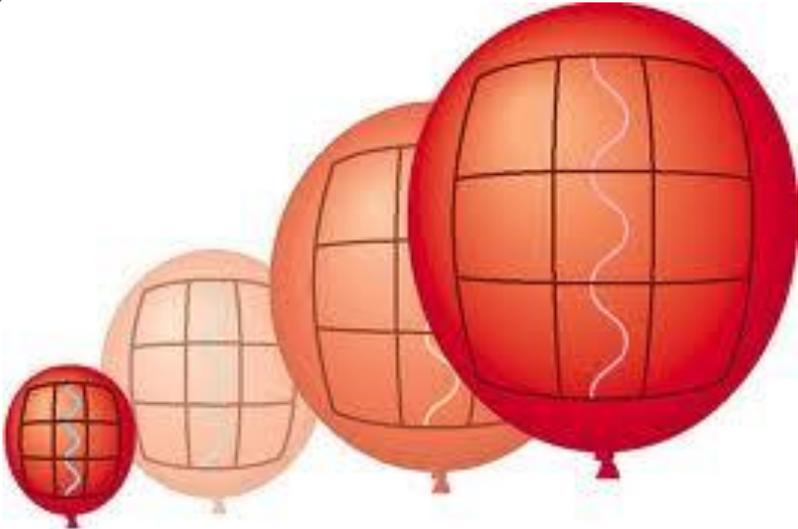




# Estadística y ciencia



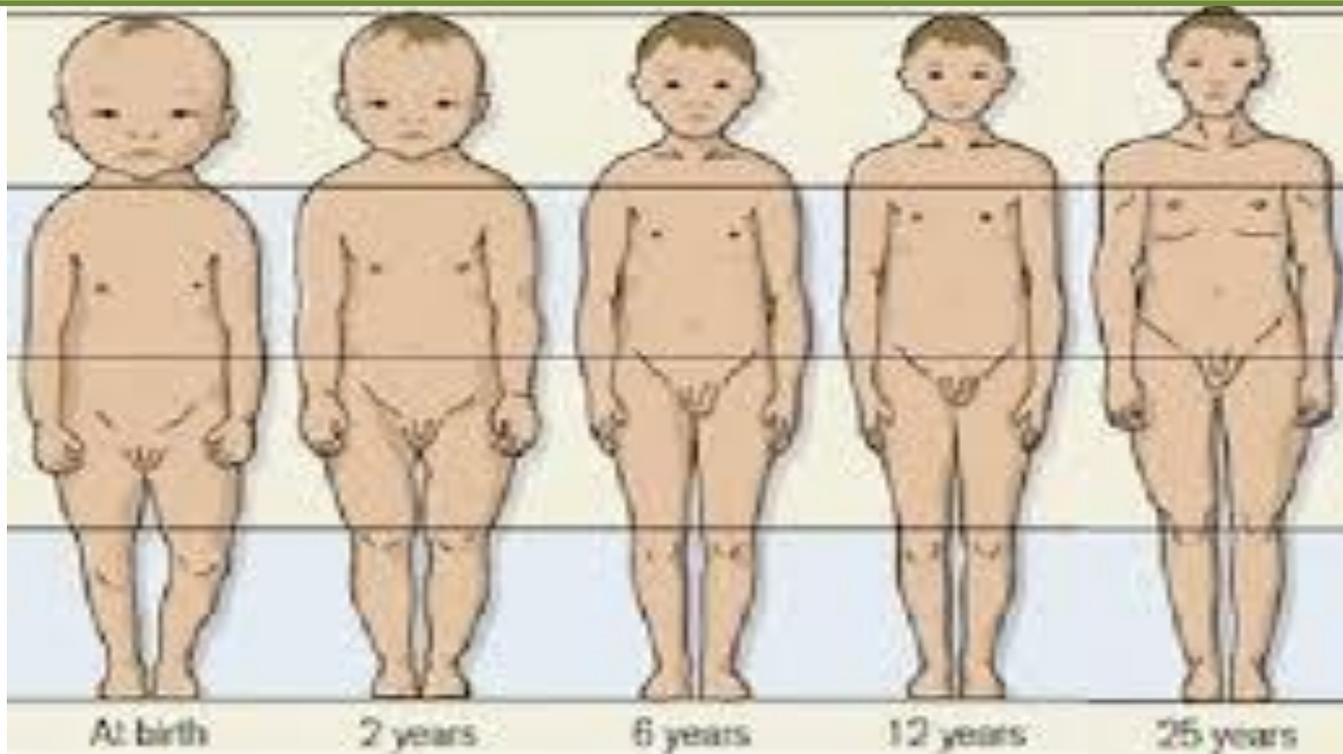
# Pensamiento proporcional





# Proporción

Relación de correspondencia entre las partes y el todo, o entre varias cosas relacionadas entre sí, en cuanto a tamaño, cantidad, dureza, etc.





# Equilibrio

situación que se da en un sistema cuando todos los factores exteriores y/o procesos internos no producen cambios .



Download from  
Dreamstime.com  
This watermark comp image is for previewing purposes only.

27721233  
Ribah2012 | Dreamstime.com





¿Una medida de la realidad?

UNA MEDIDA DE LA REALIDAD

Una estadística no significa mucho por sí sola

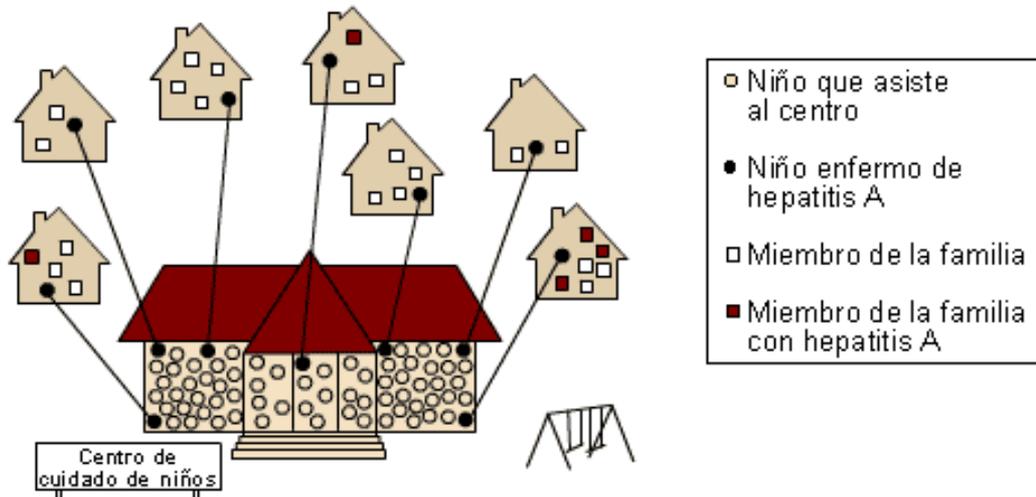
La interpretación depende del lugar tiempo y cultura de donde se observa





# Norma estadística

Figura 1  
Propagación secundaria de hepatitis A de un centro de desarrollo infantil a hogares



Tasa promedio de ocurrencia de un fenómeno





# Ideal estadístico



**Tasa de ocurrencia socialmente deseada de un fenómeno**





**Datos: información sistemáticamente adquirida que se organiza siguiendo los procedimientos de la ciencia y la estadística**

**Se conocen las limitaciones  
Del razonamiento  
y de los procesos  
matemáticos y  
se sabe el grado de  
imprecisión o de confianza**



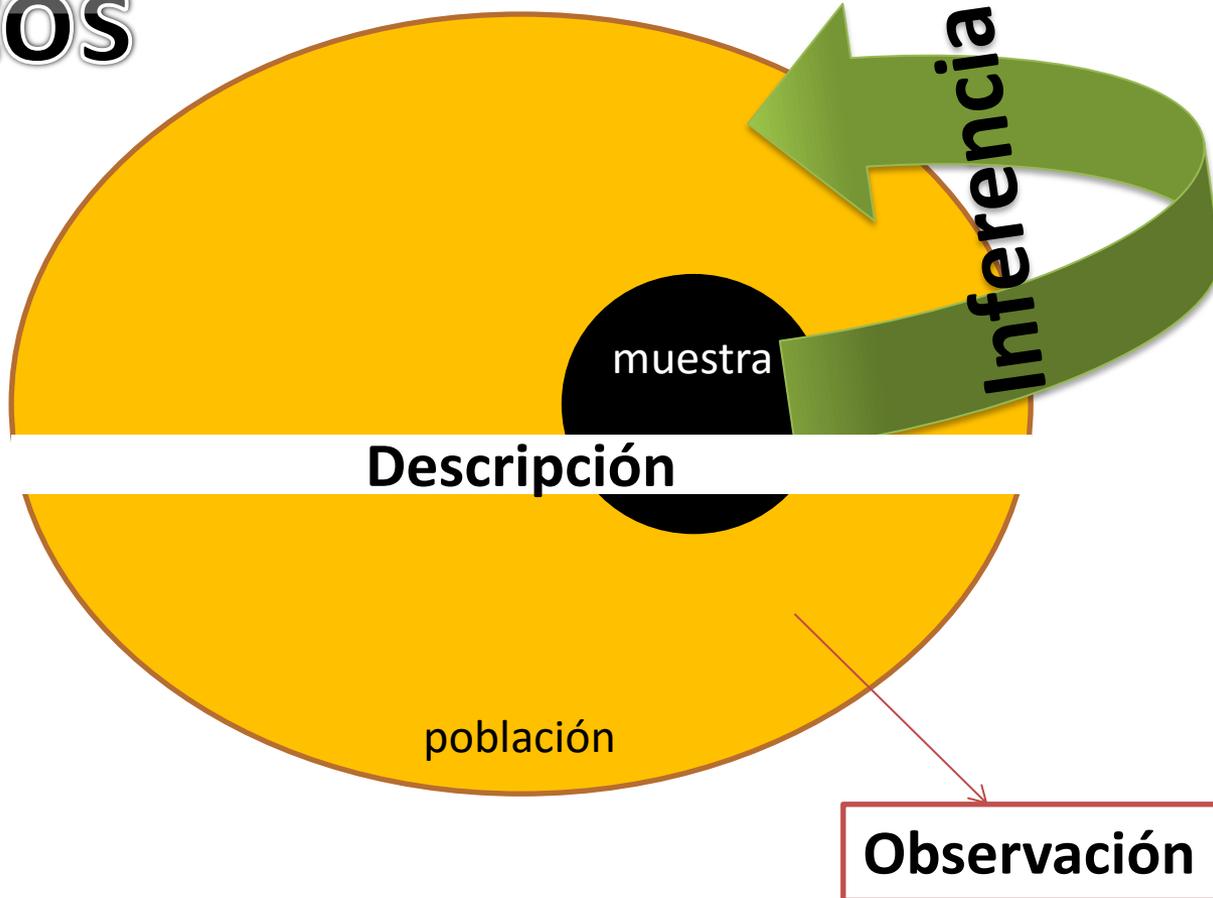
**Un objetivo es  
Controlar el  
error  
estadístico**

**Error estadístico.  
grado de imprecisión en los procedimientos utilizados  
para reunir y procesar información**



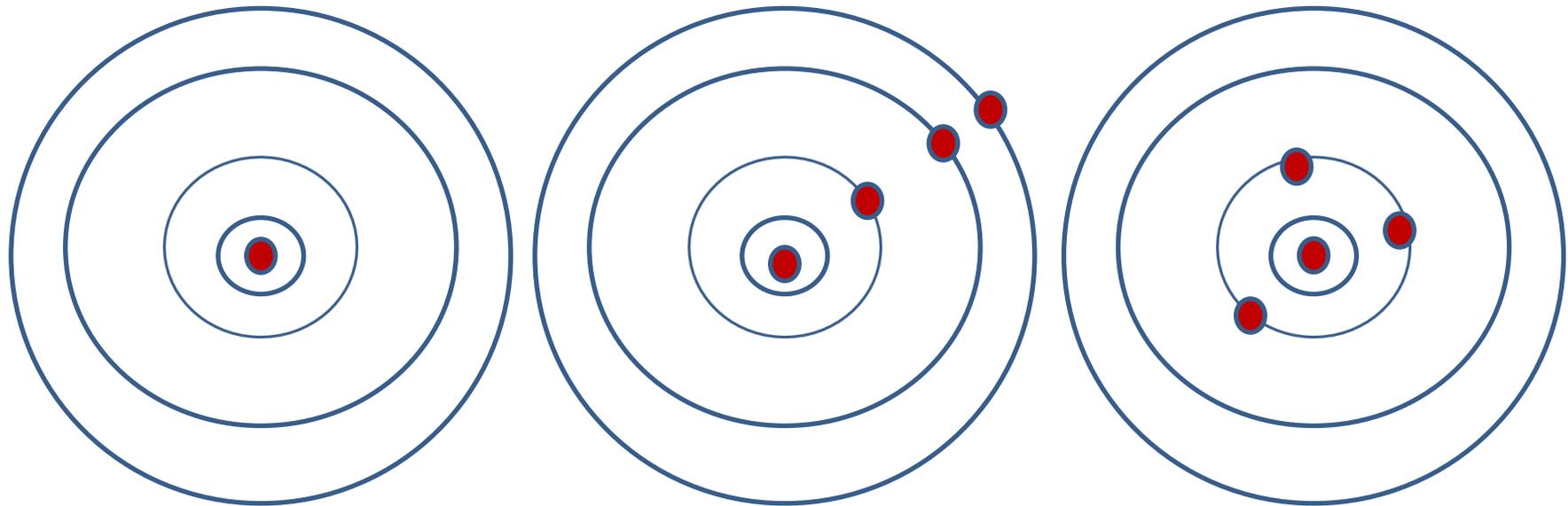


# Datos



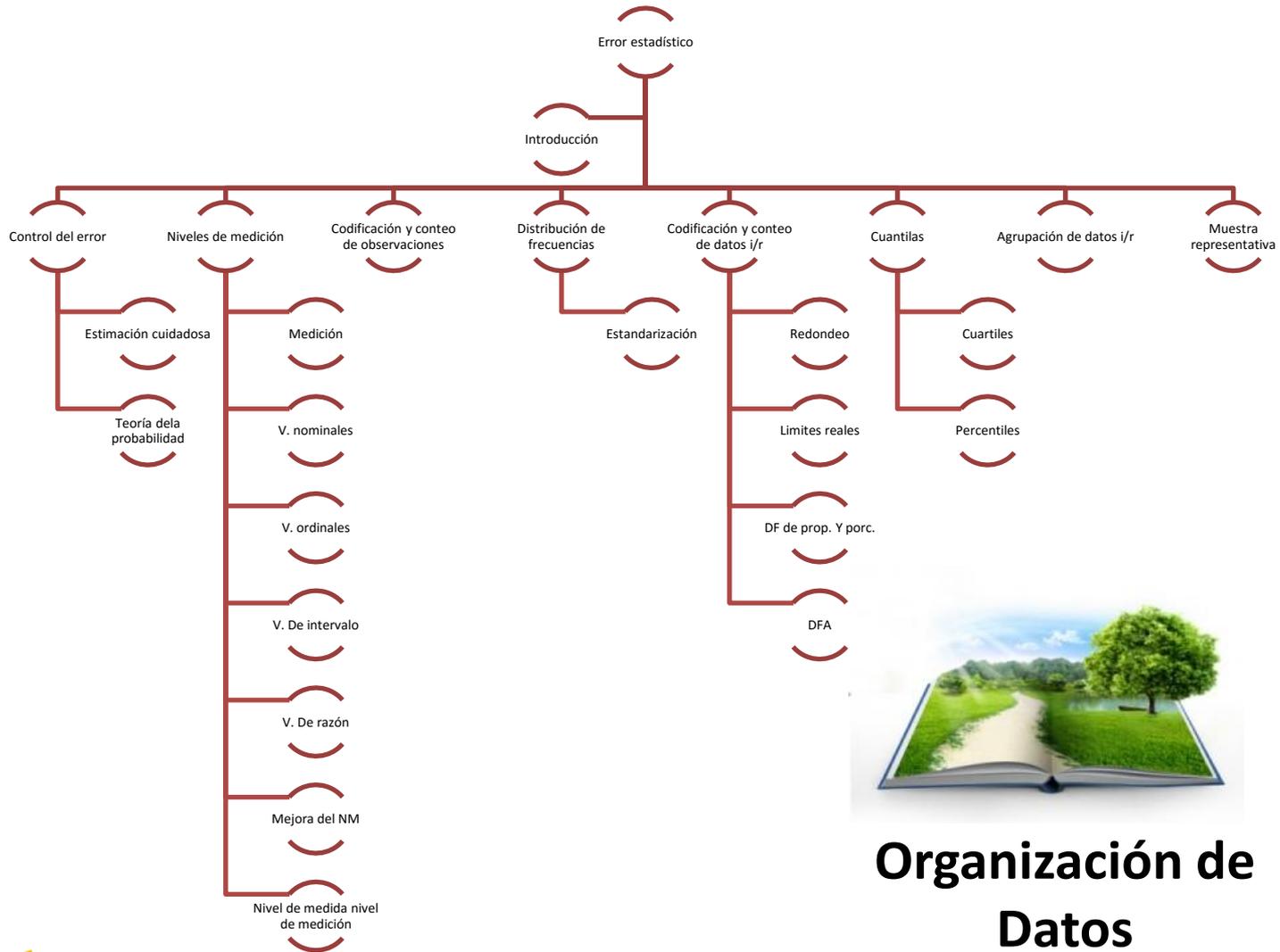


# VARIABLES, VARIACIÓN Y CONSTANTES





**Universidad Autónoma de Estado de México**  
**Facultad de Economía**  
**Maestría en Sustentabilidad, Desarrollo Regional y Metropolitano**

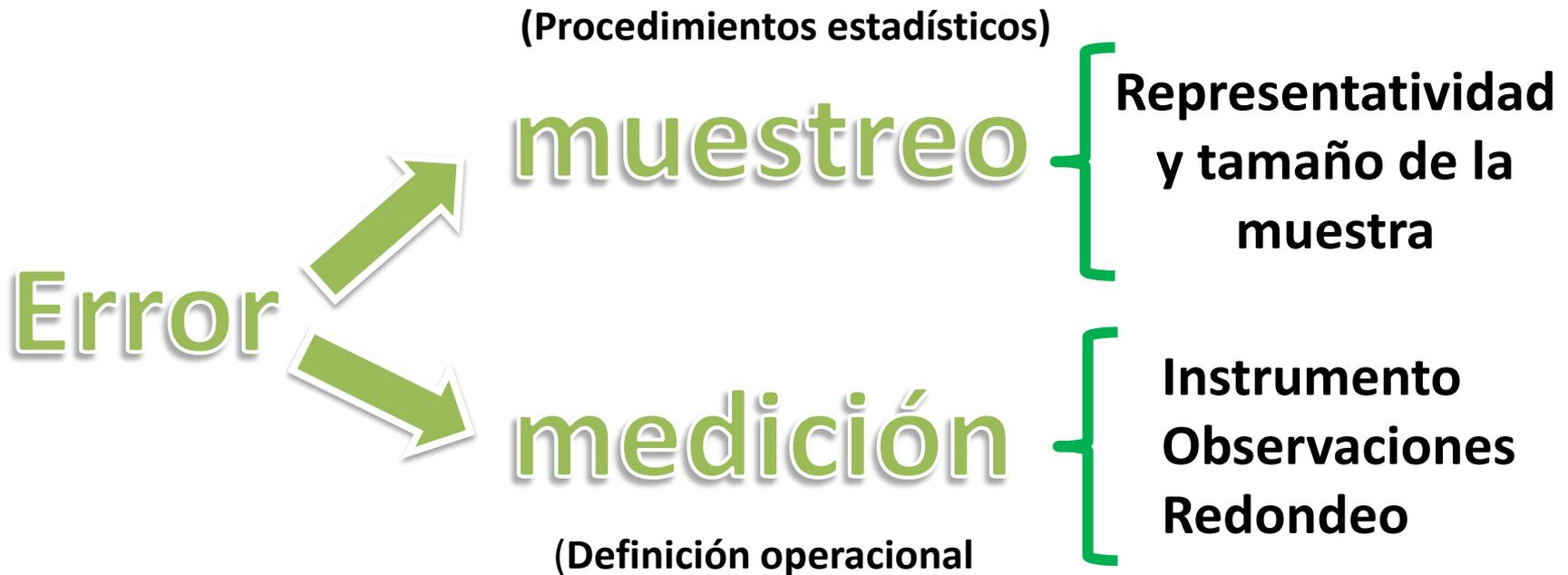


## Organización de Datos





# Fuentes potenciales de error





# Estimación cuidadosa

**Población**  
↓  
**parámetro**



**Muestra**  
↓  
**estadístico**

**Grados conocidos  
de error y confianza**

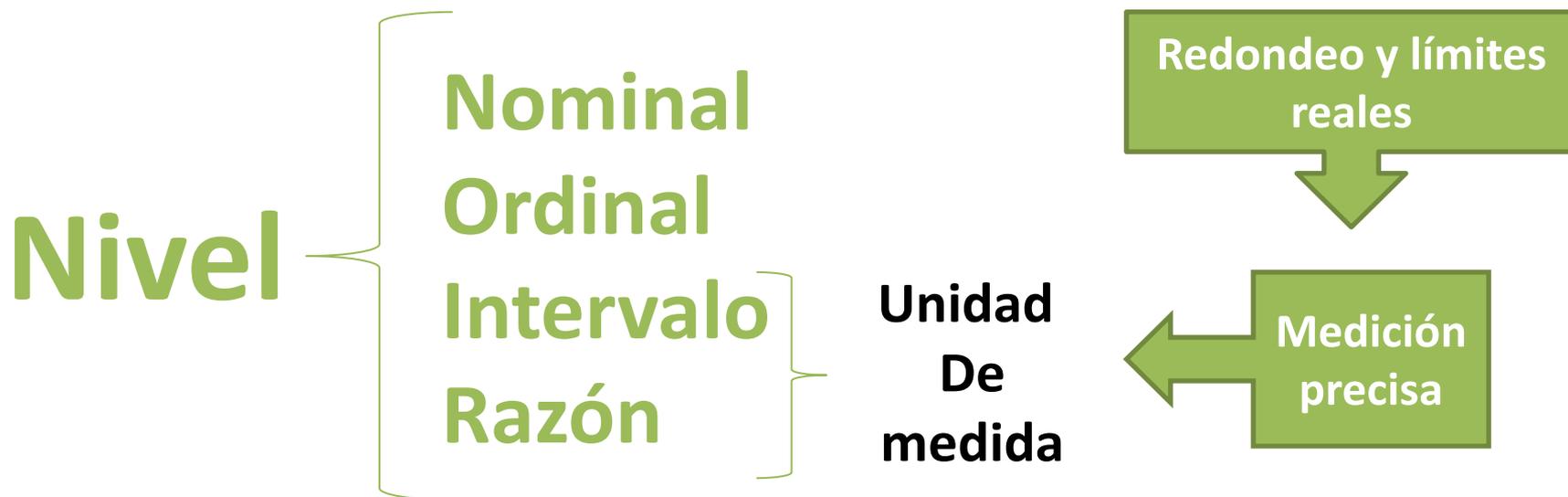
**Estimación estadística vs apresurada**





# Medición

**Símbolo que asignamos a las diferencias observadas en las cualidades o cantidades de una variable**





# Codificación y conteo

Principios:  
inclusividad y  
exclusividad



Valores perdidos

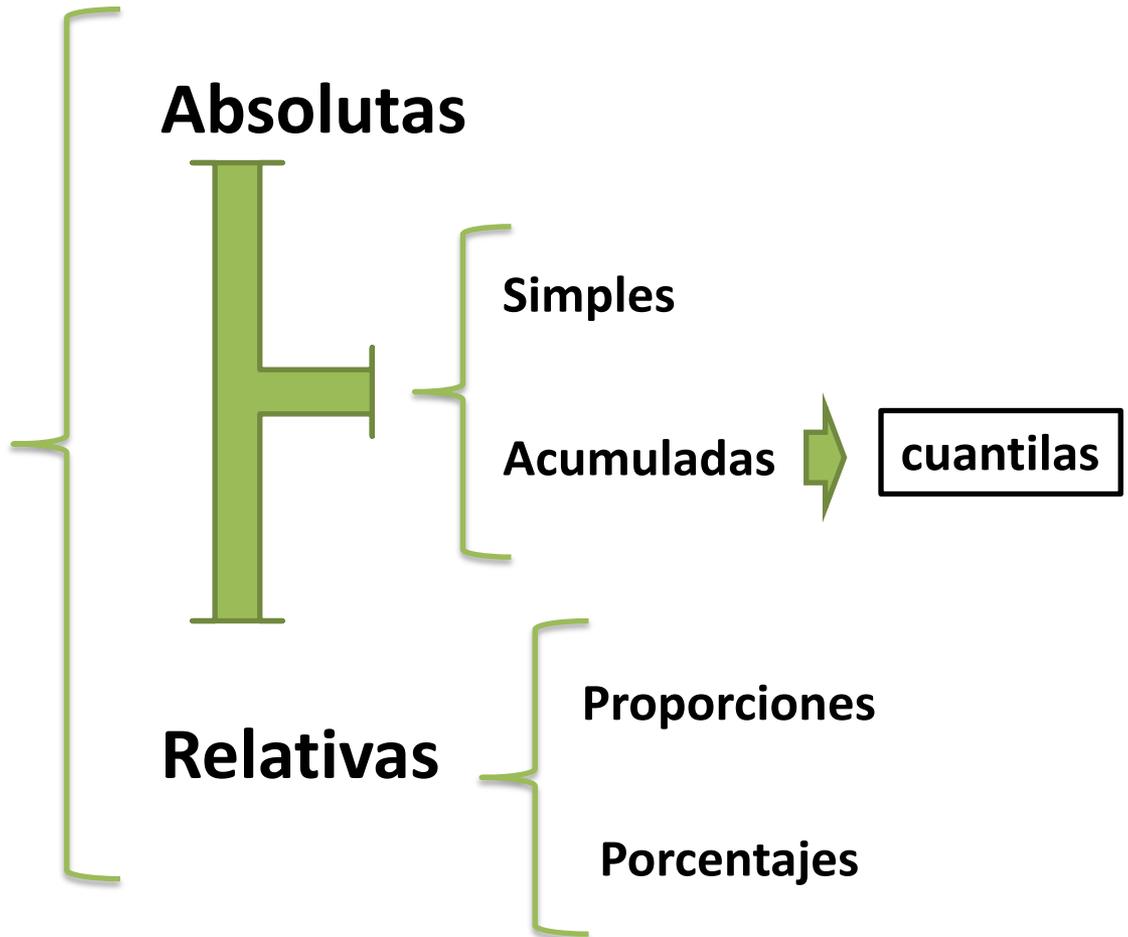
Hoja de calculo

Variable	Descripción de códigos
Edad	0=mujer, 1=hombre,9=faltante





# DISTRIBUCIONES DE FRECUENCIAS





# Series de datos

**SIMPLE**

**DE FRECUENCIAS**

**DE CLASES Y FRECUENCIAS**





Tablas y gráficas

Lineamientos

Datos nominales  
ordinales

V. Intervalo razón

Uso y aplicación

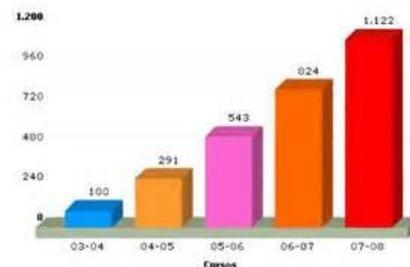
Distorsión gráfica

Gráficos de pastel

Histogramas

Gráficos de  
barras

Polígonos y líneas



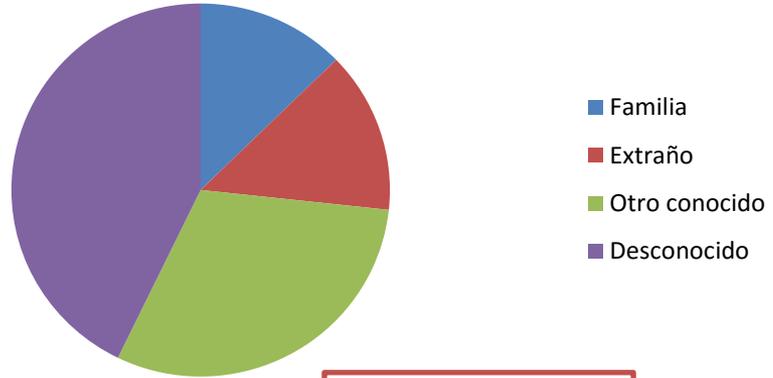
**Tablas y Gráficas**





Relación de la víctima y el delincuente	Porcentaje (%)
Familia	12.7
Extraño	14
Otro conocido	30.5
Desconocido	42.8
Total	100

Relación de la víctima y el delincuente

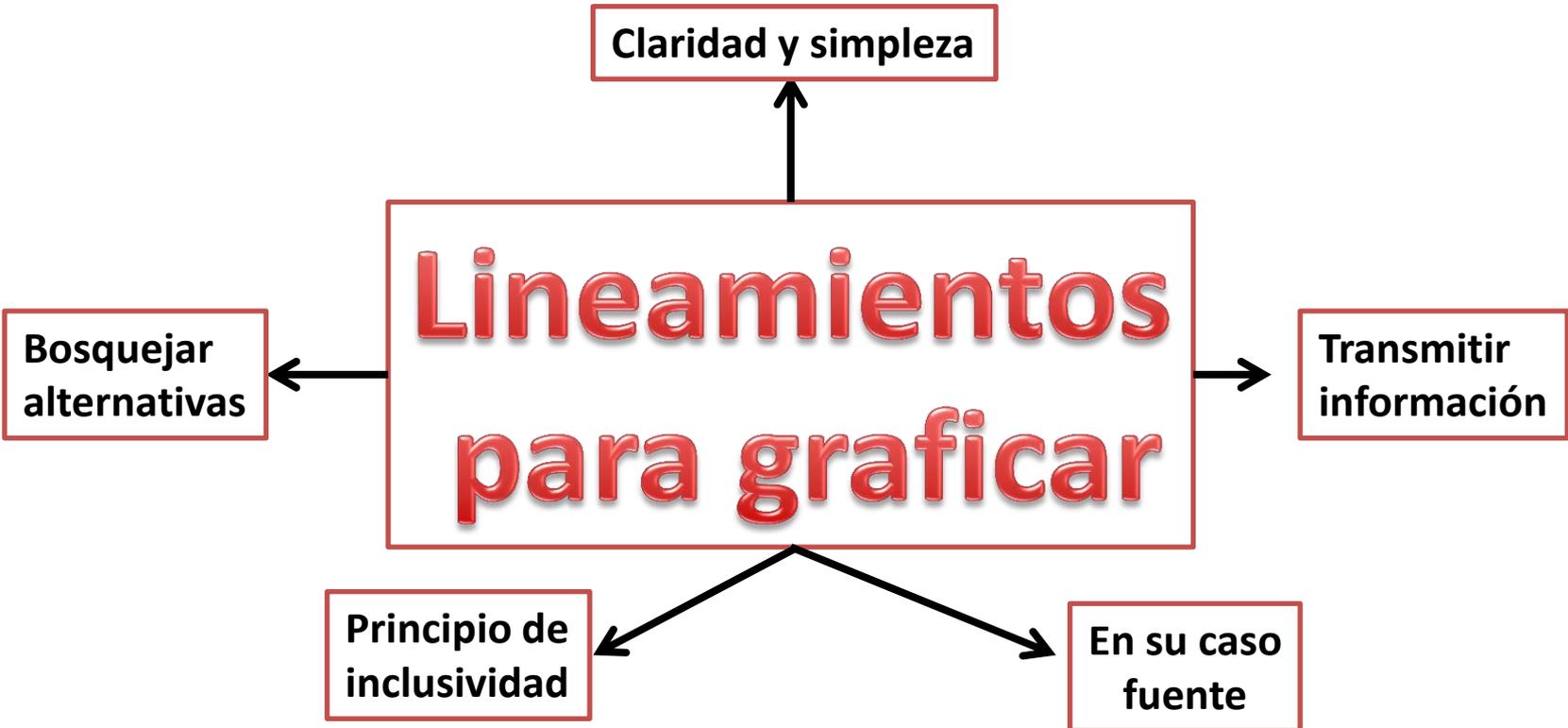


Sector/pastel

Bases para elegir Gráficos

1. Nivel de medición
2. Objetivo y relevancia
3. Características del lector



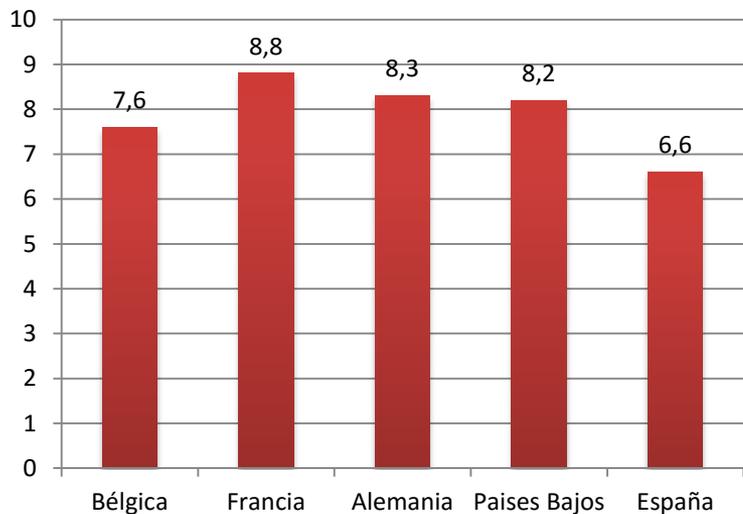




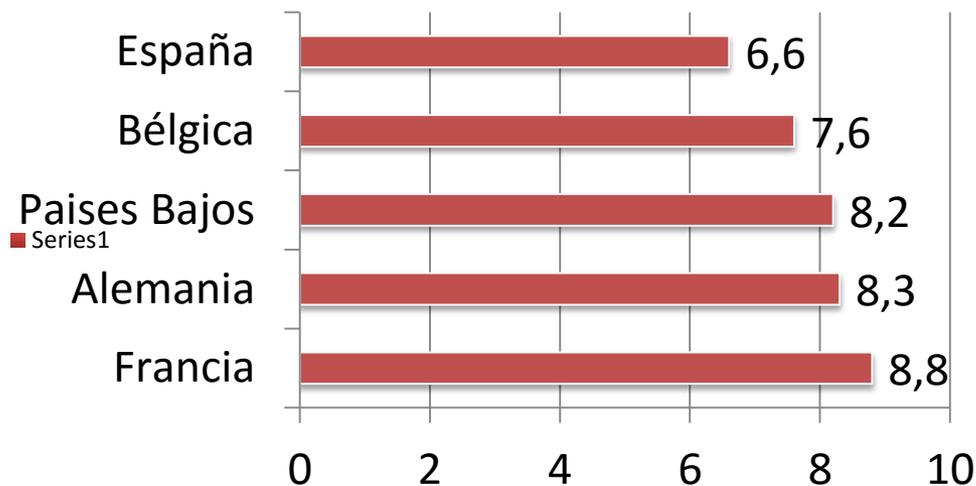
## Graficación de datos nominales/ordinales

País	% PIB gastado en atención médica en 1990
Bélgica	7.6
Francia	8.8
Alemania	8.3
Paises Bajos	8.2
España	6.6

% PIB gasto en atención médica en  
1990



% PIB gasto en atención médica en  
1990



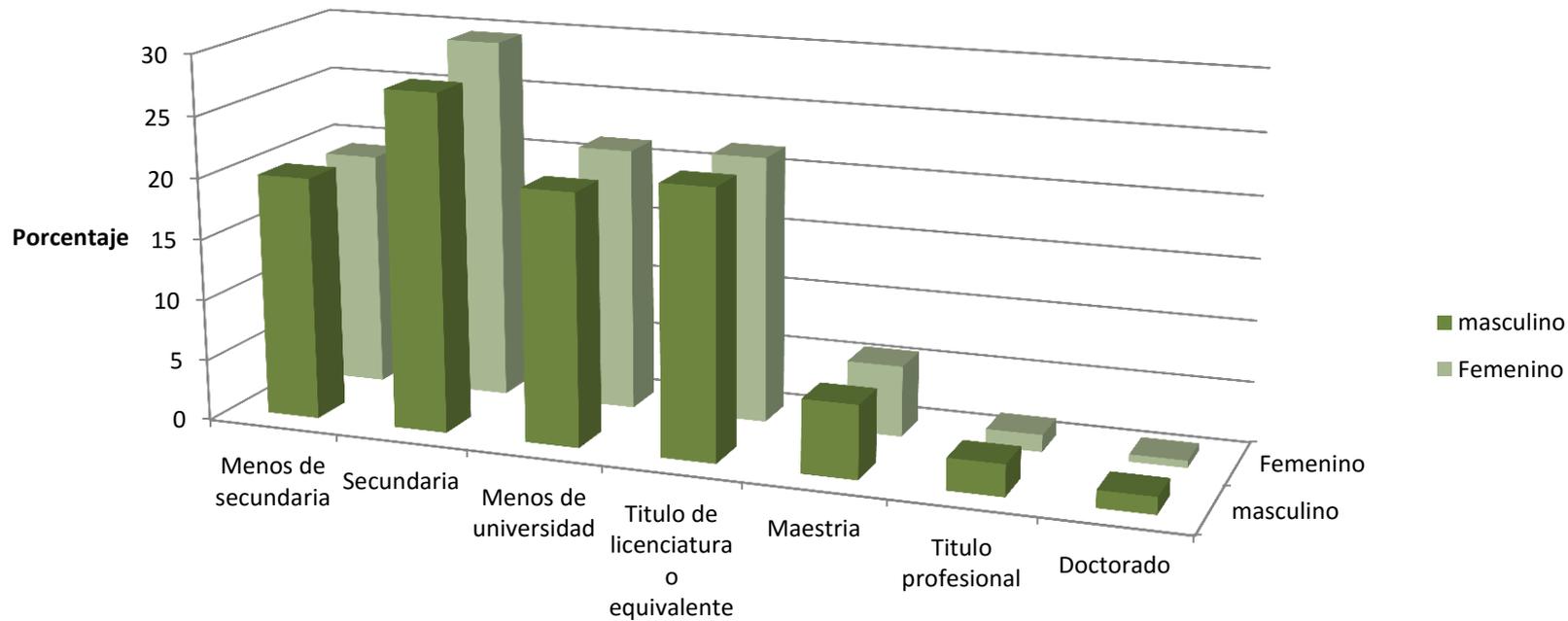
Barras





# Barras agrupadas

## Género y nivel educativo

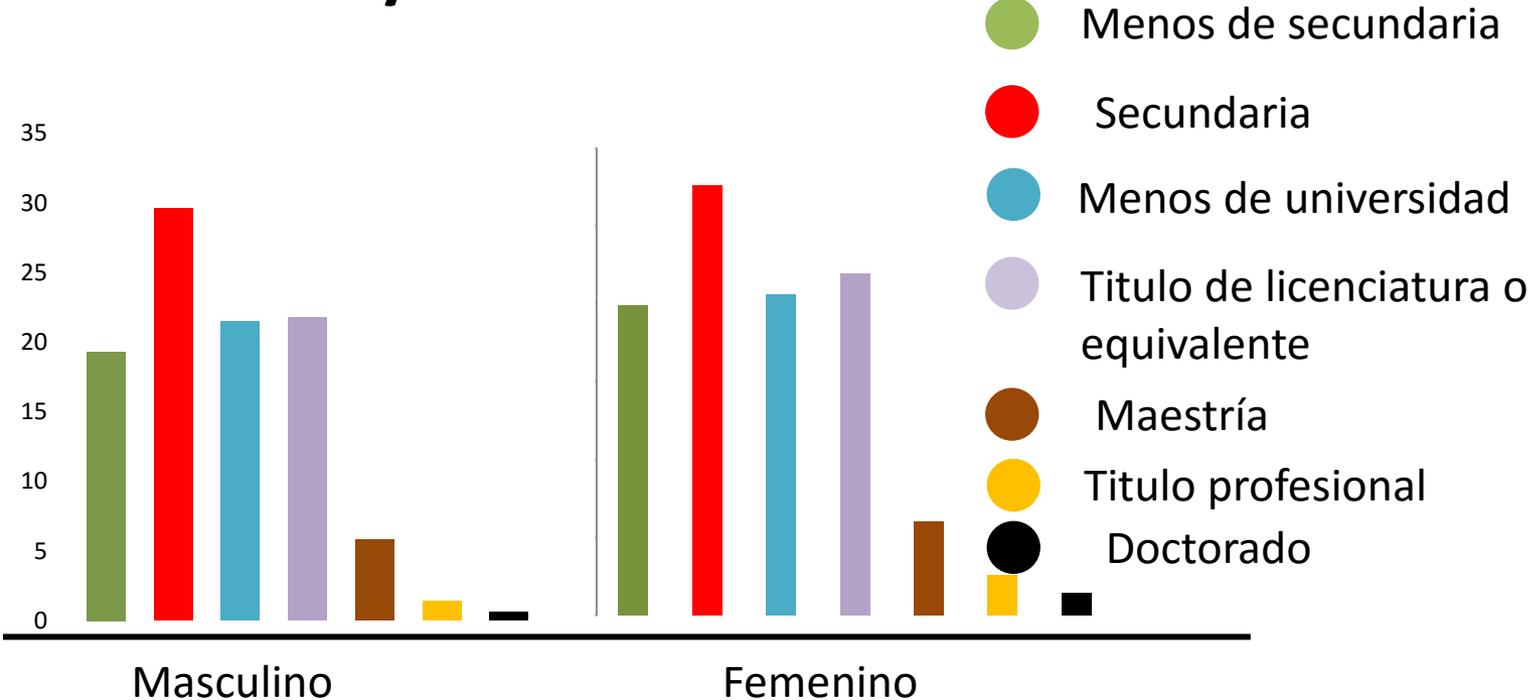


Barras agrupadas





# Género y nivel educativo



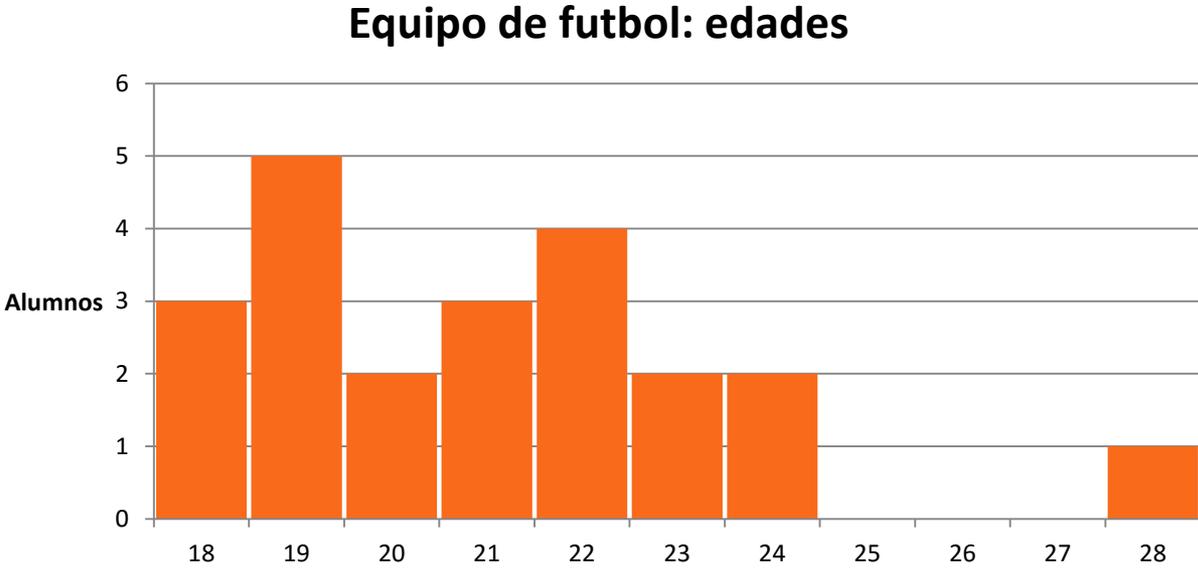
Barras agrupadas





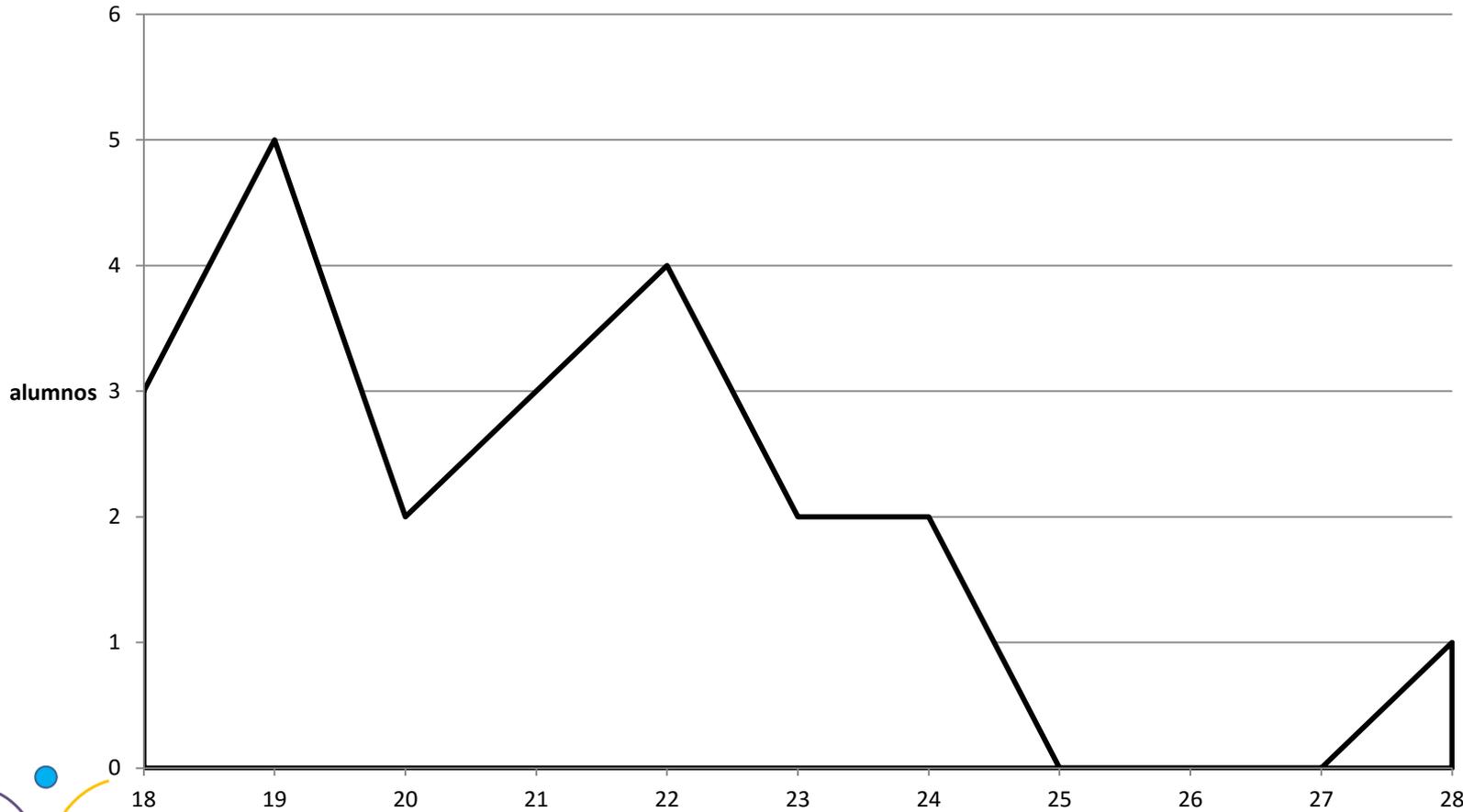
## Graficación de variables de Intervalo/razón

Edades	Estudiantes (n)
18	3
19	5
20	2
21	3
22	4
23	2
24	2
25	0
26	0
27	0
28	1
<b>suma</b>	<b>22</b>



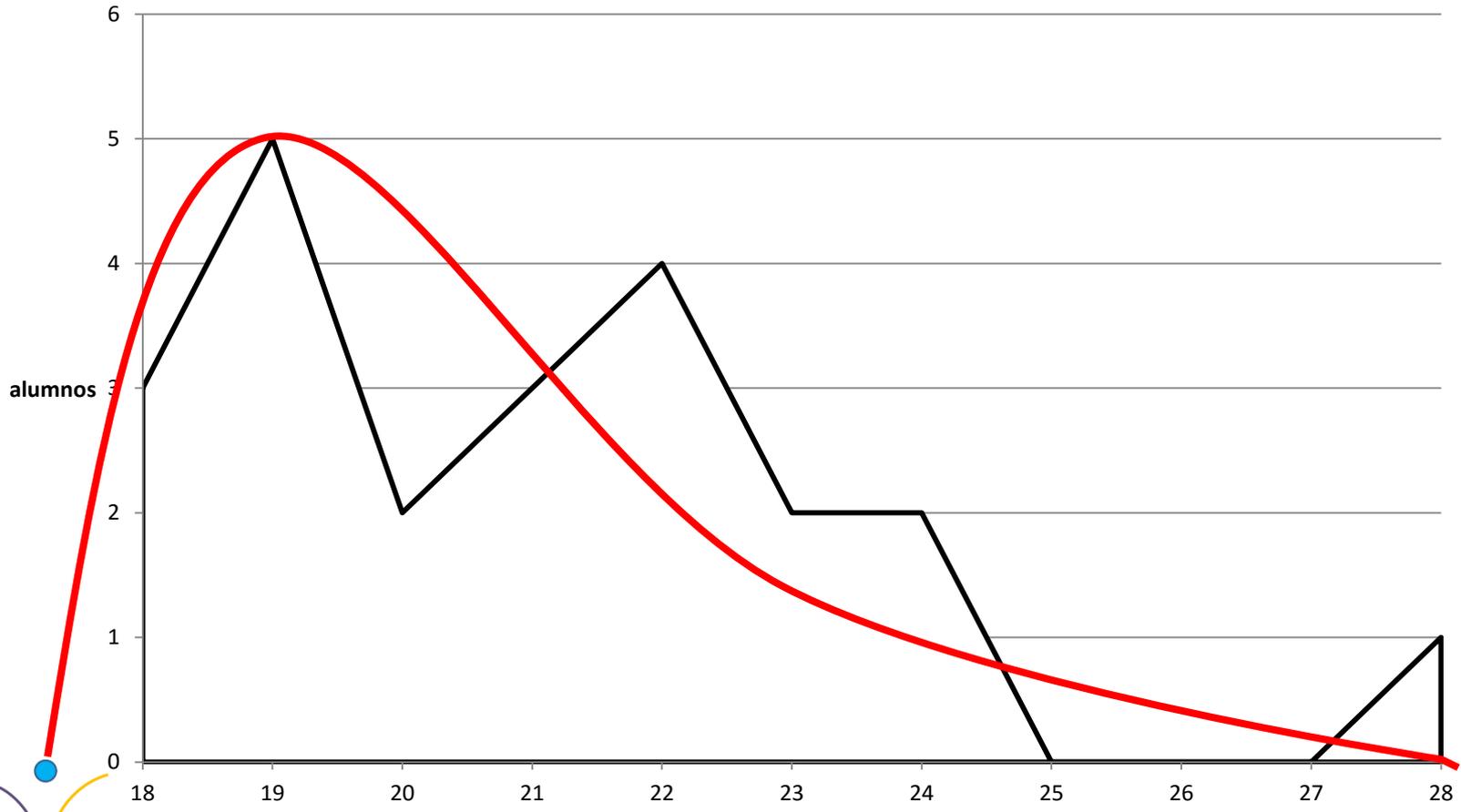


### Equipo de futbol: edades





### Equipo de futbol: edades



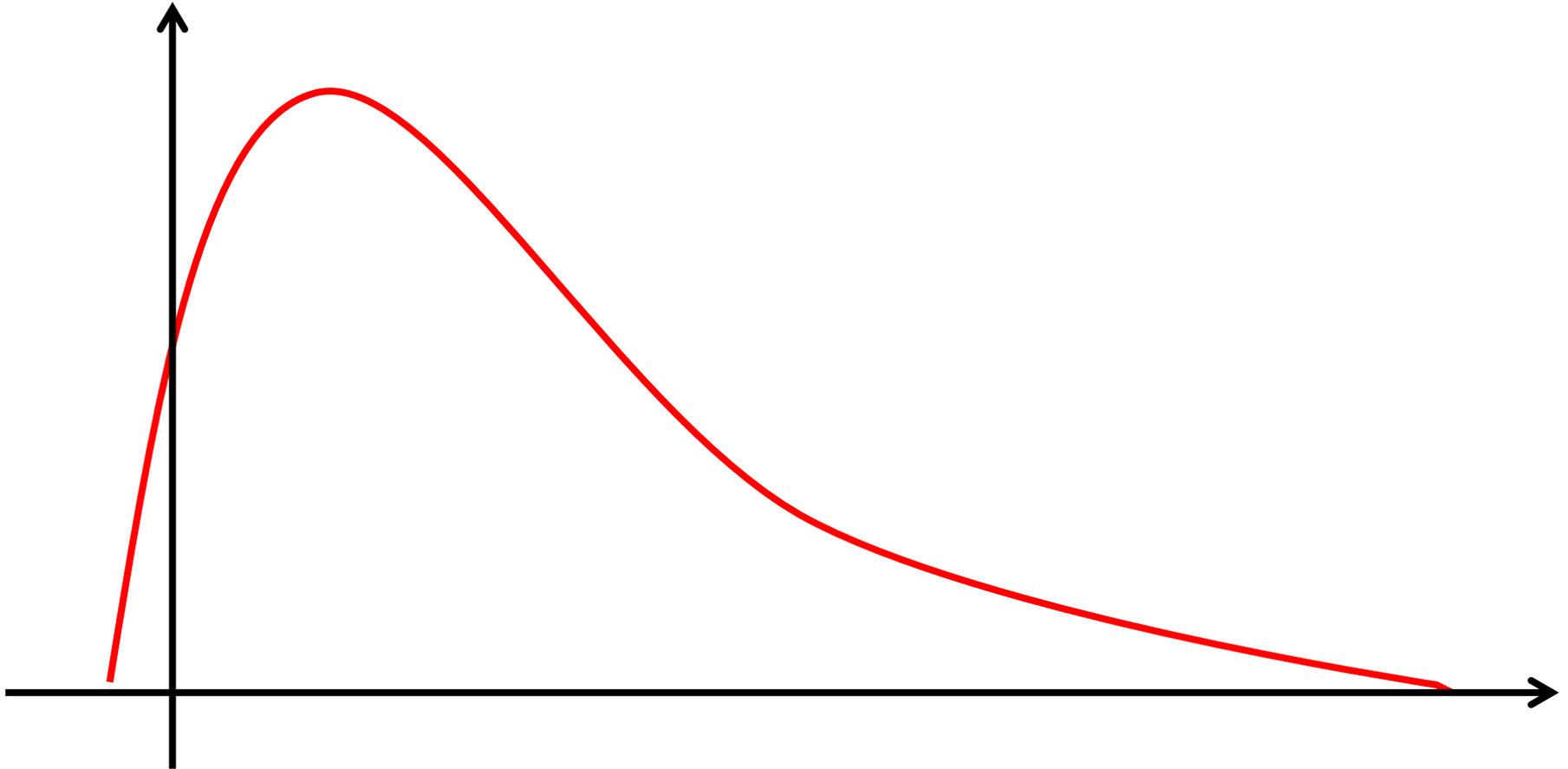


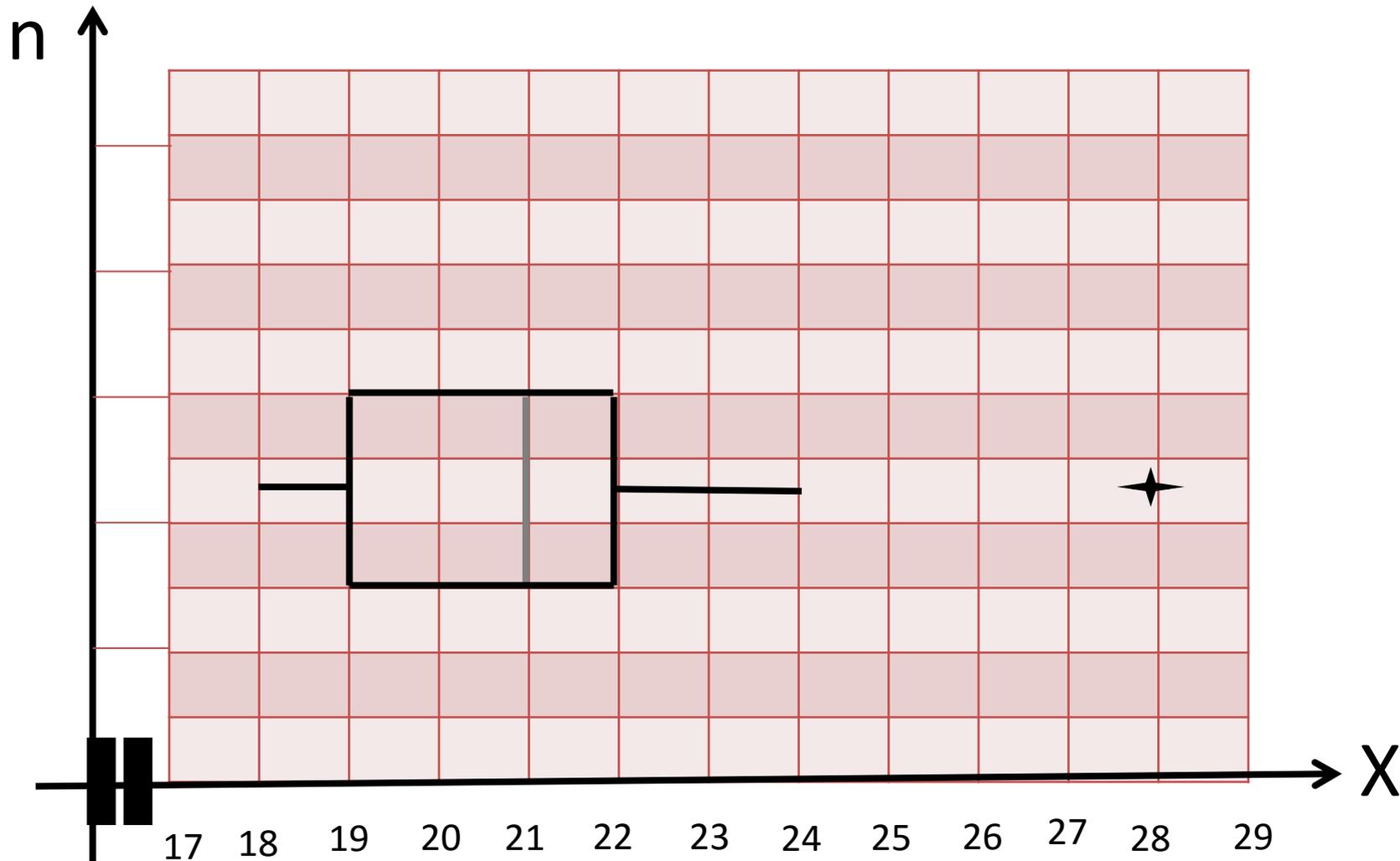


Diagrama de  
caja y bigote

# Resumen de 5 números

Edades	Estudiantes (n)	f	F
18	3	13.6	13.6
19	5	22.7	36.4
20	2	9.1	45.5
21	3	13.6	59.1
22	4	18.2	77.3
23	2	9.1	86.4
24	2	9.1	95.5
25	0	0.0	95.5
26	0	0.0	95.5
27	0	0.0	95.5
28	1	4.5	100.0
<b>suma</b>	<b>22</b>	<b>100</b>	

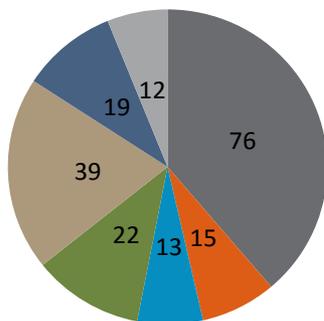
<b>Min</b>	$P_{25}$	$P_{50}$	$P_{75}$	<b>Max</b>
<b>18</b>	<b>19</b>	<b>21</b>	<b>22</b>	<b>28</b>



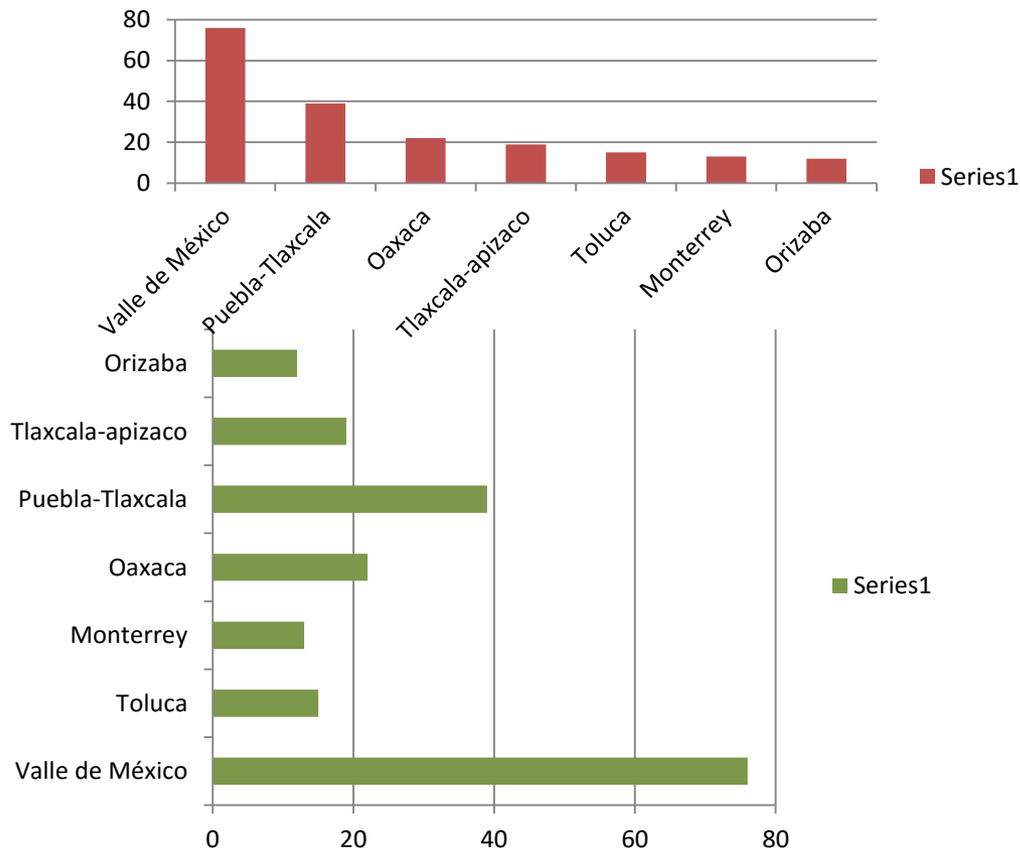


### Mexico: Número de Municipios de las Zonas Metropolitanas con más de 10 municipios, 2010

- Valle de México
- Toluca
- Monterrey
- Oaxaca
- Puebla-Tlaxcala
- Tlaxcala-apizaco
- Orizaba

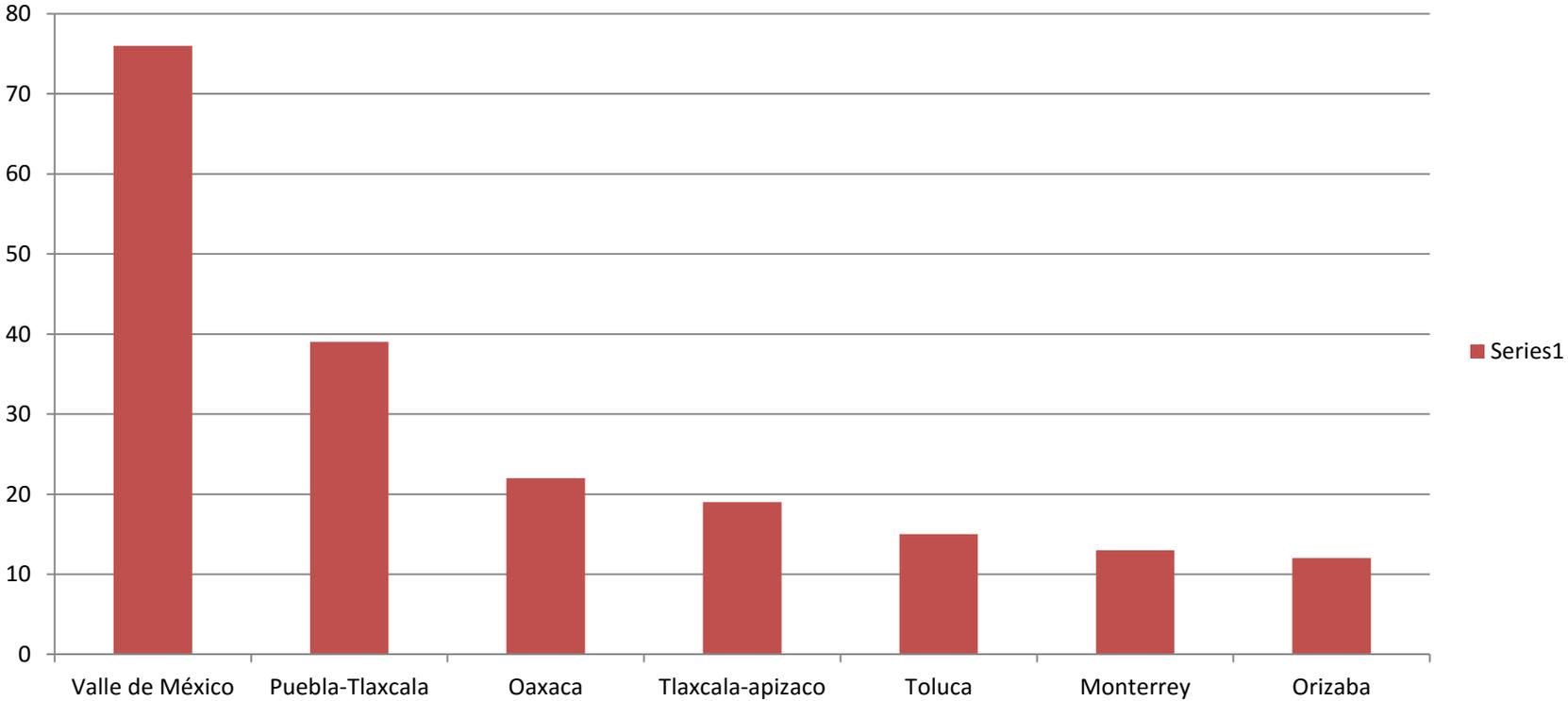


### México: Número de municipios de las Zonas Metropolitanas con mas de 10 municipios, 2010



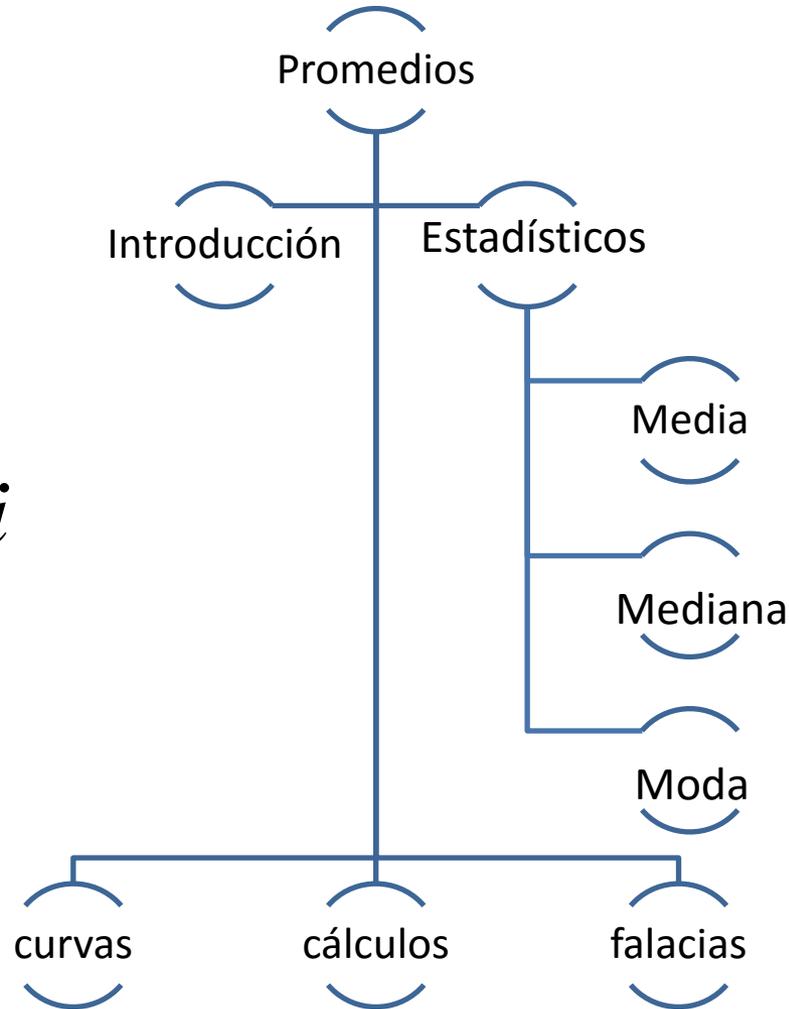


### México: Número de municipios de las Zonas Metropolitanas con mas de 10 municipios, 2010



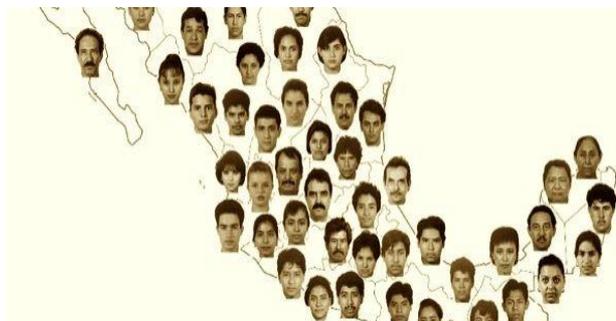


$$\sum_{i=1}^n x_i$$





# ¿Cuánto mide México?



Rangos de edad	Estatura		Peso	
	H	M	H	M
>18	1.64	1.58	74.8	68.7
18-25	1.67	1.61	70.4	62.9
40-50			77.3	72.2

Datos de 17 mil 364 personas mayores de 18 años. [Cámara Nacional de la Industria del Vestido](http://www.muyinteresante.com.mx/preguntas-y-respuestas/12/02/09/medidas-poblacion-mexicana/)

<http://www.muyinteresante.com.mx/preguntas-y-respuestas/12/02/09/medidas-poblacion-mexicana/>





Una medida de posición o de tendencia es un número que se toma como **orientación** para referirnos a un conjunto de datos.

La media aritmética muestral representa el **centro físico del conjunto de datos** y se define como la suma de los valores observados, dividido por el total de observaciones.

**Serie simple**

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

**Serie de frecuencias**

$$\bar{x} = \frac{\sum_{c=1}^{\infty} f_c x_c}{\sum_{c=1}^{\infty} f_c}$$

**Serie de clases y frecuencias**

$$\bar{x} = \frac{\sum_{c=1}^{\infty} pm_c x_c}{\sum_{c=1}^{\infty} pm_c}$$





Propiedades  
de la Media  
aritmética

La suma de las desviaciones respecto de la media es igual a cero,

$$\sum_{i=1}^n (x_i - \bar{x}) = 0$$

La suma de los cuadrados de las desviaciones a partir de la media aritmética es menor que la suma de los cuadrados de las desviaciones a partir de cualquier otro valor.

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

La media no se altera por una transformación lineal de escala.

$$a + b\bar{x}$$





## Sentido de proporcionalidad

### Ventajas

1. Es fácil de calcular
2. Se entiende fácilmente.
3. Se presta a operaciones algebraicas.
4. Es el estadístico más comúnmente usado

### Desventajas

Está desproporcionalmente afectada por los valores extremos y que, por consiguiente, puede estar muy lejos de ser una representación de la muestra, por lo que no es recomendable usarla en distribuciones muy asimétricas.

$$\bar{x}_1 + \bar{x}_2 = \frac{\sum_{i=1}^{n_1} x_1 + \sum_{i=1}^{n_2} x_2}{n_1 + n_2}$$

$$\text{Pero si } n_1 = n_2 \rightarrow \frac{\bar{x}_1 + \bar{x}_2}{2}$$





La mediana muestral representa el **valor central de los datos** cuando las mediciones están ordenadas

**Serie simple**                      Posición  $\tilde{x} = 0.5(n + 1)$   
 $\tilde{x}$  = valor de la Posición  $\tilde{x}$

**Serie de frecuencias**                      Posición  $\tilde{x} = 0.5 \sum f_i$   
 $\tilde{x}$  = valor de la Posición  $\tilde{x}$

**Serie de clases y  
frecuencias**

$$\tilde{x} = L_{\text{inf}} + \left[ \frac{0.5(\sum f_i) - F}{f_i} \right] * c$$





## Desventajas

Es insensible a los valores de la variable X, pero es muy sensible con el tamaño de muestra n

Indicador	Unidad de medida	Año	Valor
Población total <sup>a</sup>	Miles de habitantes	2010	112 337
Tiempo de duplicación <sup>a</sup>	Años	2010	39.5
Tasa de crecimiento media anual de la población <sup>a</sup>	Por ciento	2005-2010	1.8
Densidad de población <sup>a</sup>	Habitantes/km <sup>2</sup>	2010	57
<b>Edad mediana <sup>a</sup></b>	<b>Años</b>	<b>2010</b>	<b>26</b>

<http://www3.inegi.org.mx/sistemas/temas/default.aspx?s=est&c=17484>





La moda representa el valor con la mayor **frecuencia de los datos.**

**Serie simple**  $\hat{x} = \text{Valor que más se repite}$

**Serie de frecuencias**  $\hat{x} = \text{Valor con la frecuencia más alta}$

**Serie de clases y frecuencias**

$$\hat{x} = L_{\text{inf}} + \left[ \frac{\Delta_1}{\Delta_1 + \Delta_2} \right] * c \therefore$$

$$\Delta_1 = f - f_{\text{ant}} \text{ y } \Delta_2 = f - f_{\text{post}}$$





## Relación entre Media Aritmética, Mediana y Moda:

Para distribuciones unimodales que sean poco asimétricas:

$$\bar{X} - \hat{X} = 3(\bar{X} - \tilde{X})$$

Sus posiciones relativas, según la simetría de la distribución de frecuencias es:

Relación	Simetría
$\hat{X} = \tilde{X} = \bar{X}$	Simétrica
$\hat{X} < \tilde{X} < \bar{X}$	Sesgo positivo
$\hat{X} > \tilde{X} > \bar{X}$	Sesgo negativo





## Otras medidas de tendencia central

### Media Aritmética Ponderada

Es el promedio de los datos en donde se le da un peso o importancia específica a cada observación. Se calcula:

$$\bar{x}_w = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}$$

### Media Geométrica

sirve para promediar los crecimientos geométricos de una variable.

Si suponemos que Y representa el factor de crecimiento geométrico de la variable X, es decir:

$$y_i = \frac{x_i}{x_{i-1}}$$





## Otras medidas de tendencia central

Entonces el factor de crecimiento geométrico promedio de la variable X será:

**Serie simple**

$$G = \sqrt[n]{\prod_{i=1}^n y_i}$$

**Serie de  
frecuencias**

$$G = \sqrt[n]{\prod_{i=1}^k y_i^{f_i}}$$

**Media Armónica**

Se usa cuando los datos a promediarse están medidos en unidades expresadas en forma de cocientes (km./hr., \$/lt, etc.),





## Serie simple

$$H = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

## Serie de frecuencias

$$H = \frac{n}{\sum_{i=1}^k \frac{f_i}{x_i}}$$





## MEDIDAS DE DISPERSIÓN

**Rango** (o Intervalo): Es la distancia que existe entre el menor y mayor valor de los datos.

$$R = \max - \min$$

**Rango Semi-Inter Cuartil (Q): (o Desviación Cuartil)**

Mide el rango promedio de una cuarta parte de los datos (evita los valores extremos)

$$Q = \frac{Q_3 - Q_1}{2}$$

**Desviación Media Absoluta (DM):** Es la distancia promedio de los datos a su media.

$$DM = \frac{\sum_{i=1}^n |X_i - \bar{X}|}{n}$$

$$DM = \frac{\sum_{i=1}^k f_i |X_i - \bar{X}|}{n}$$





**Varianza muestral** : La suma de las distancias al cuadrado se divide entre en número de datos menos uno:

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

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

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

$$S^2 = \left( \frac{\sum_{i=1}^k f_i x_i^2}{n - 1} \right) - \left( \frac{n\bar{x}^2}{n - 1} \right)$$





### Varianza poblacional:

Es el promedio del cuadrado de la distancia de los datos a su media

$$\sigma^2 = \frac{\sum_{i=1}^N (X_i - \mu)^2}{N}$$

$$\sigma^2 = \left( \frac{\sum_{i=1}^N X_i^2}{N} \right) - \mu^2$$

$$\sigma^2 = \frac{\sum_{i=1}^k f_i (X_i - \mu)^2}{N}$$

$$\sigma^2 = \left( \frac{\sum_{i=1}^k f_i * X_i^2}{N} \right) - \mu^2$$

### Desviación Estándar:

Mide la variación de los datos en términos absolutos. Es la raíz cuadrada positiva de la varianza.

$$S = \sqrt{S^2}$$

$$\sigma = \sqrt{\sigma^2}$$





La desviación estándar se interpreta construyendo intervalos alrededor del promedio:

**a) Teorema de Chebyshev.**

Si la distribución no es simétrica y unimodal o se desconoce su forma:

- ✓ **Al menos el 75%** de los valores cae dentro de **2 desviaciones** estándar alrededor de la media.
- ✓ **Al menos el 89%** de los valores caen dentro de **3 desviaciones** estándar alrededor de la media.





## b) Regla Empírica.

Si la distribución es una curva acampanada, unimodal y simétrica:

- ✓ Aproximadamente **el 68%** de los datos (población) se encuentran a una **desviación** estándar alrededor de la media.
- ✓ Aproximadamente **el 95%** de los datos (población) se encuentran a **2 desviaciones** estándar alrededor de la media.
- ✓ Aproximadamente **el 99%** de los datos (población) se encuentran a **3 desviaciones** estándar alrededor de la media:





**Coeficiente de Variación (cv):** Mide la variación relativa de la variable con respecto a su promedio. Mide la magnitud de la desviación estándar en relación con la magnitud de la media. Se expresa en por cientos.

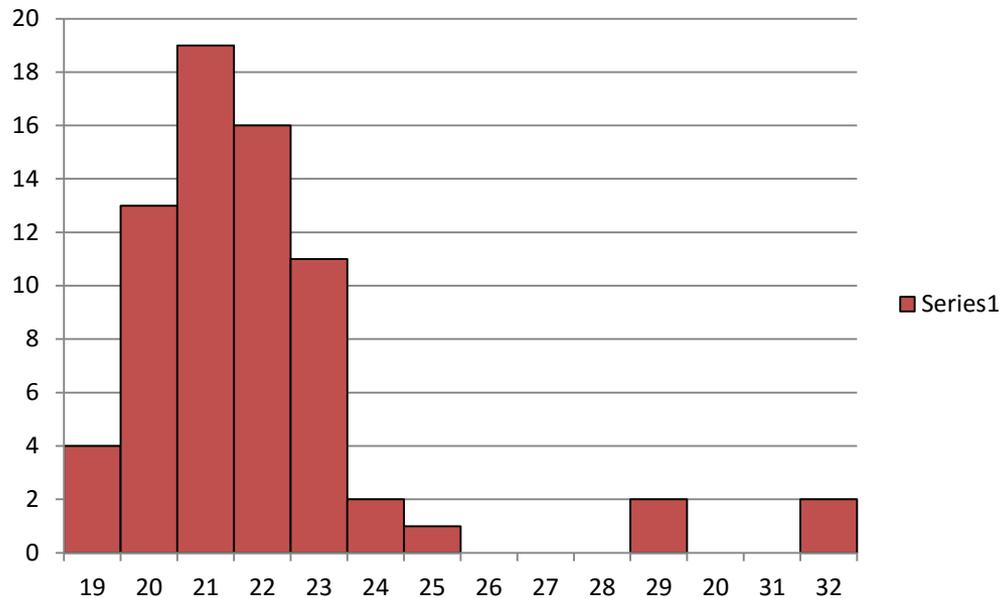
$$CV = \frac{S}{\bar{X}} \times 100$$





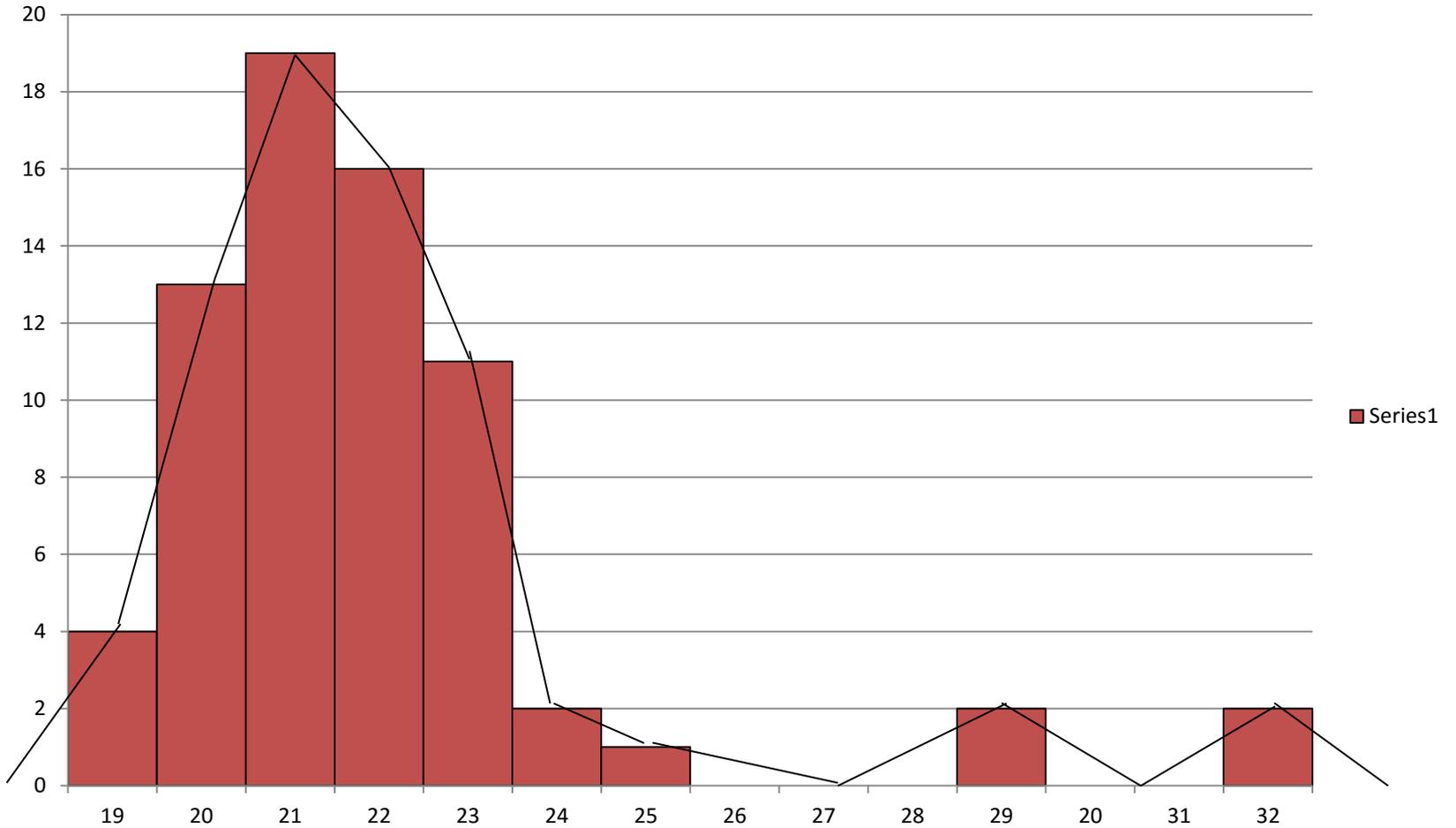
**Universidad Autónoma de Estado de México**  
**Facultad de Economía**  
**Maestría en Sustentabilidad, Desarrollo Regional y Metropolitano**

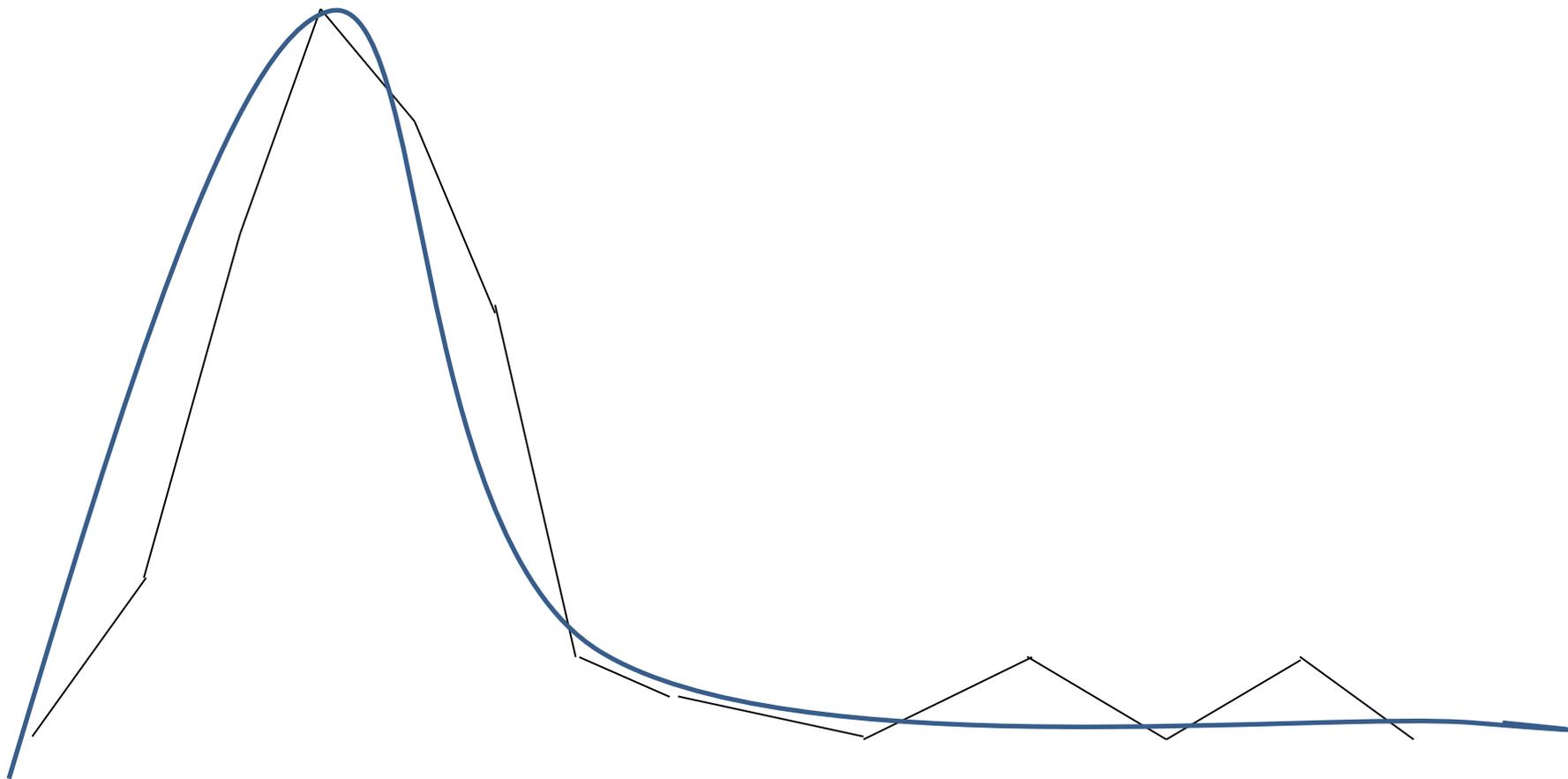
Edades	f	f%	F%
19	4	5.7	5.7
20	13	18.6	24.3
21	19	27.1	51.4
22	16	22.9	74.3
23	11	15.7	90.0
24	2	2.9	92.9
25	1	1.4	94.3
26	0	0.0	94.3
27	0	0.0	94.3
28	0	0.0	94.3
29	2	2.9	97.1
20	0	0.0	97.1
31	0	0.0	97.1
32	2	2.9	100.0
	70	100.0	





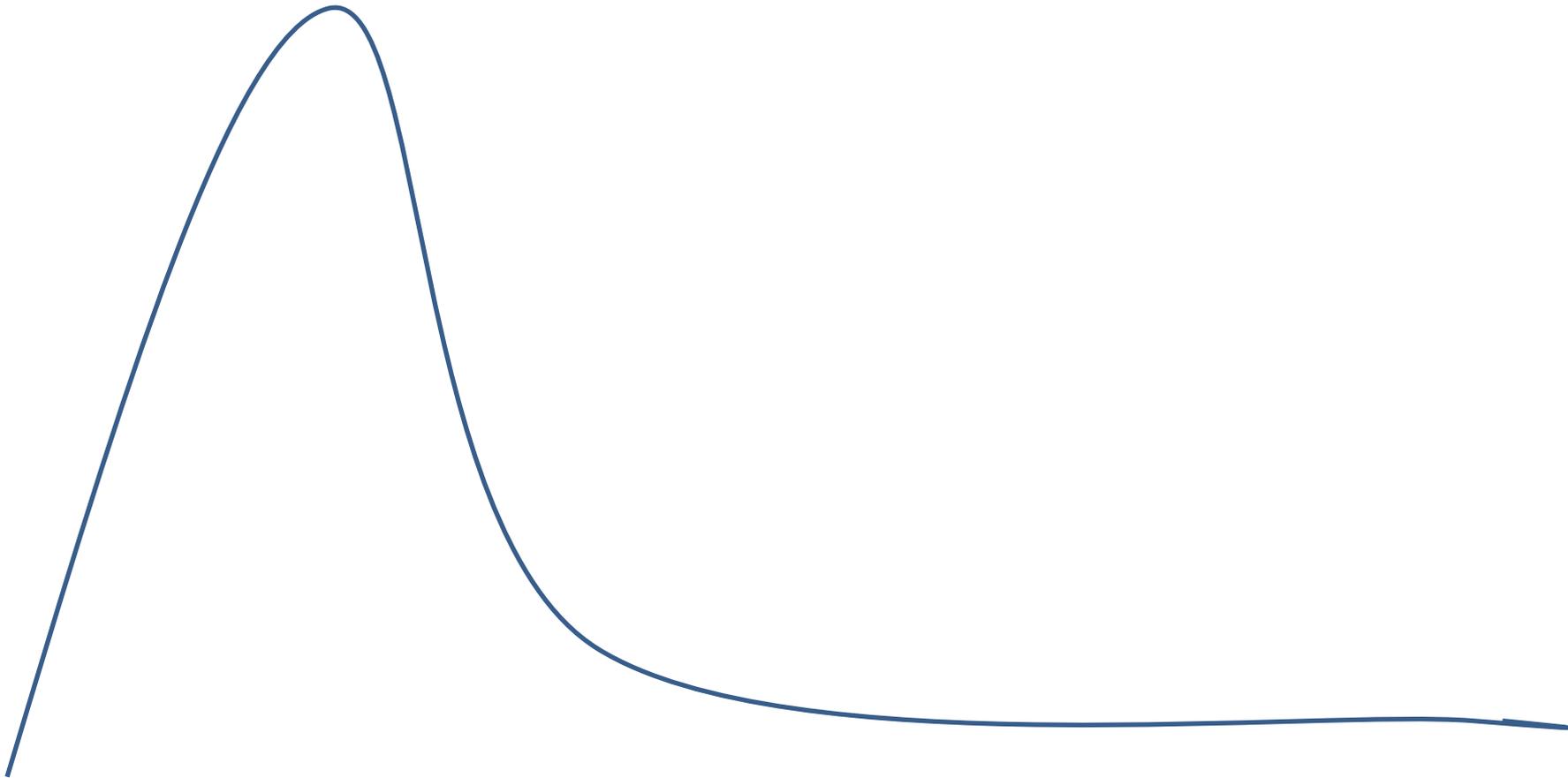
Universidad Autónoma de Estado de México  
Facultad de Economía  
Maestría en Sustentabilidad, Desarrollo Regional y Metropolitano







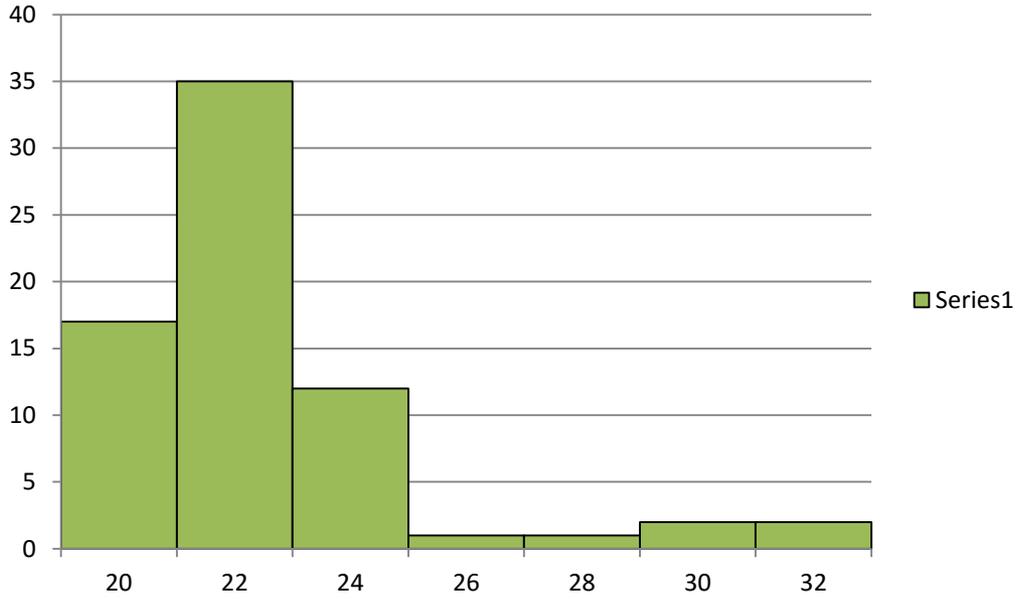
Universidad Autónoma de Estado de México  
Facultad de Economía  
Maestría en Sustentabilidad, Desarrollo Regional y Metropolitano





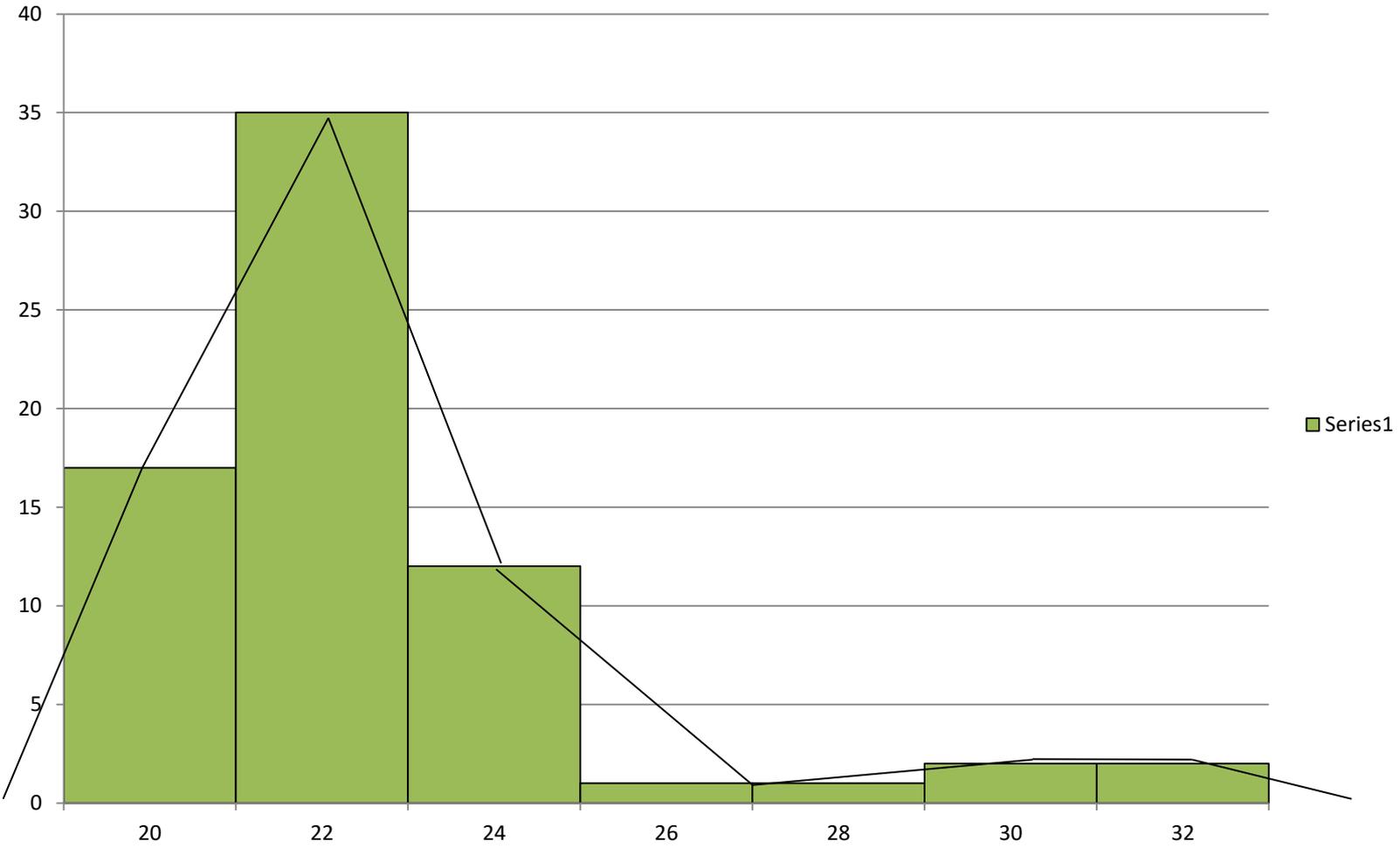
Universidad Autónoma de Estado de México  
Facultad de Economía  
Maestría en Sustentabilidad, Desarrollo Regional y Metropolitano

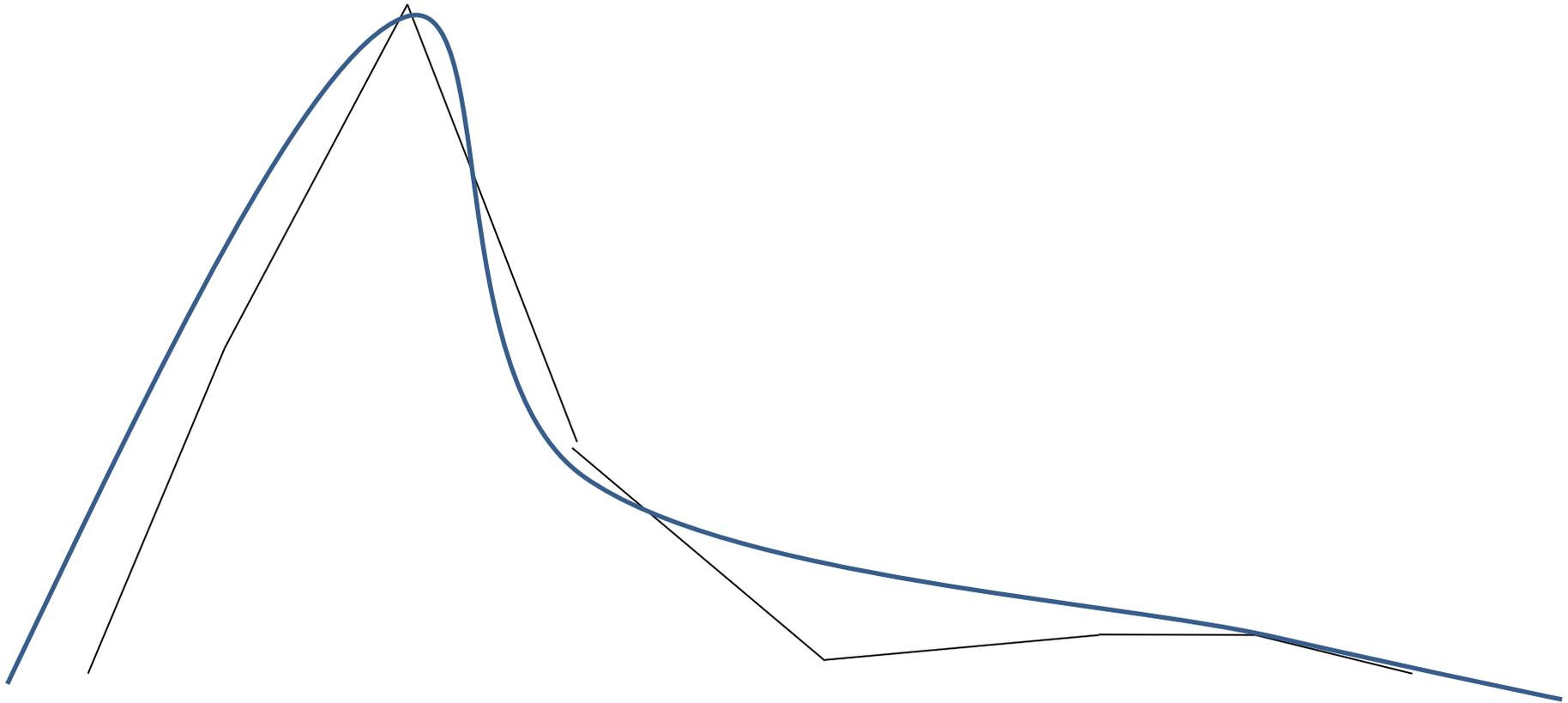
LI	LS	PM	f	f%
19	21	20	17	24.29
21	23	22	35	50.00
23	25	24	12	17.14
25	29	27	2	2.86
29	31	30	2	2.86
31	33	32	2	2.86
			70	100.00

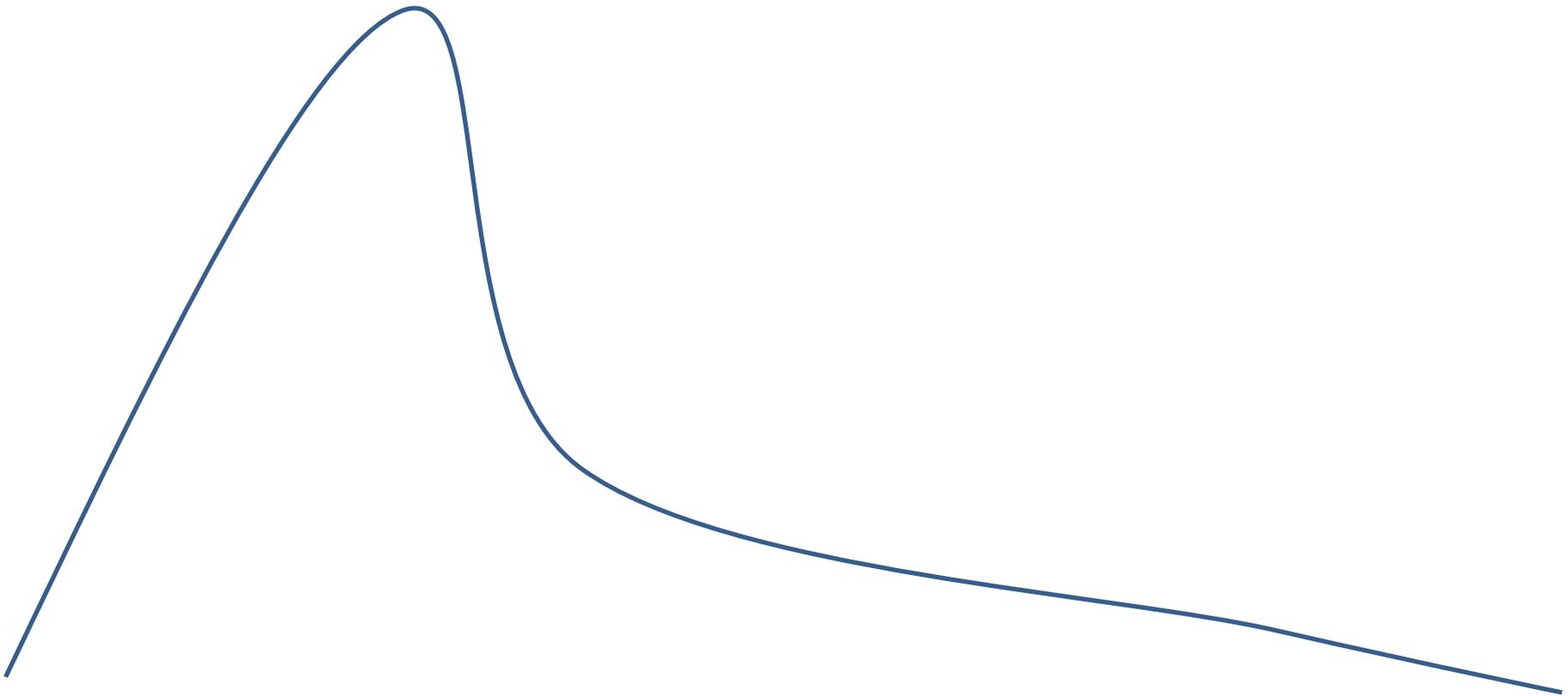




Universidad Autónoma de Estado de México  
Facultad de Economía  
Maestría en Sustentabilidad, Desarrollo Regional y Metropolitano

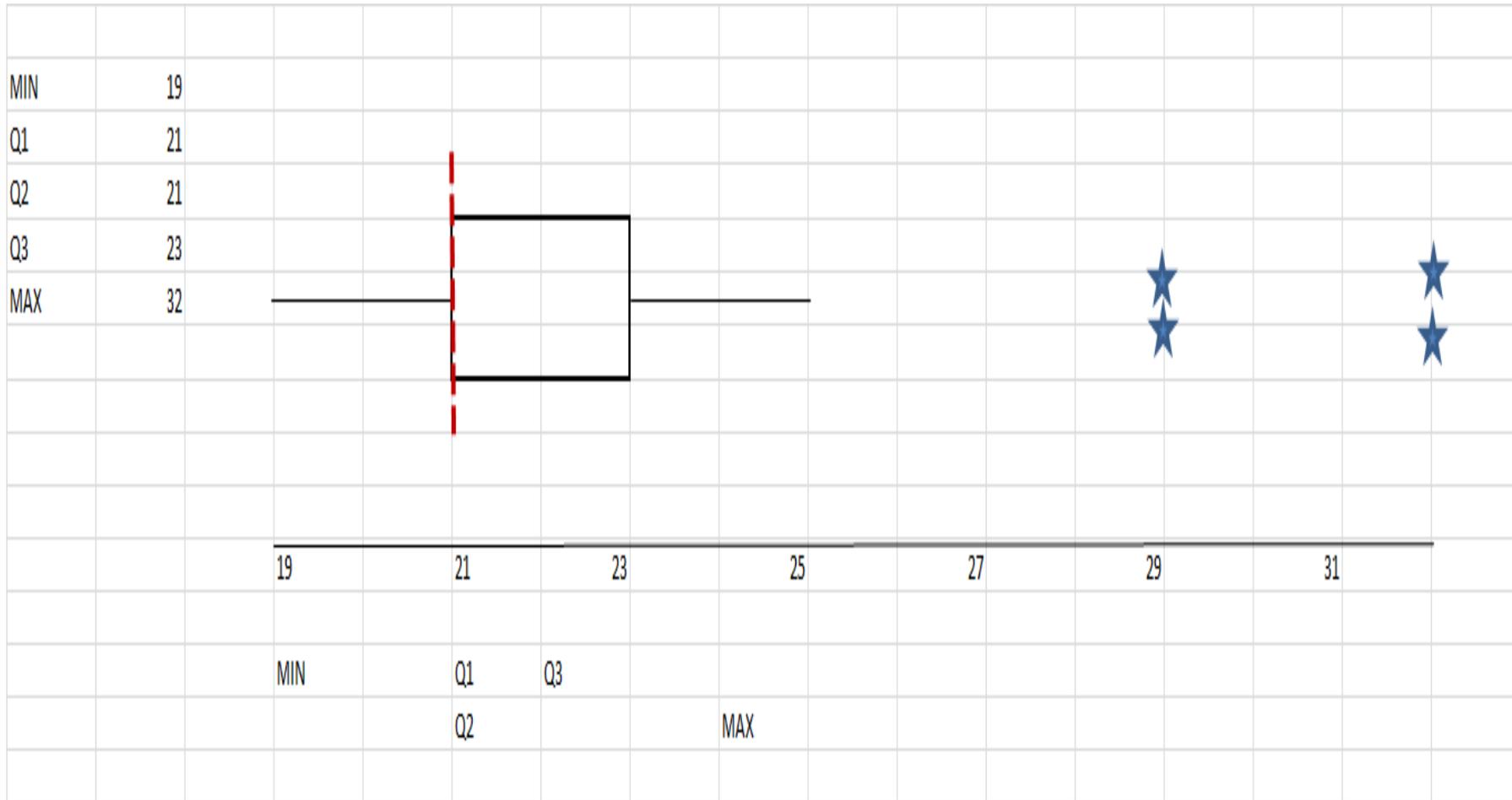








Universidad Autónoma de Estado de México  
Facultad de Economía  
Maestría en Sustentabilidad, Desarrollo Regional y Metropolitano





Edades	
Media	22
Error típico	0
Mediana	21
Moda	21
Desviación estándar	3
Varianza de la muestra	6
Curtosis	7
Coefficiente de asimetría	2
Rango	13
Mínimo	19
Máximo	32
Suma	1535
Cuenta	70
CV	11





## Bibliografía



Wackerly, D.D., W. Mendenhall y R. L. Scheaffer (2002),  
***Estadística matemática con aplicaciones***, México,  
Thomson.

Mendenhall, W., R.J. Beaver y B. Beaver (2007)  
***Introducción a la probabilidad y estadística. Internacional***  
Thomson Editores S. A. de C. V.

Freund, J.E., I. Miller y M. Miller (2000), ***Estadística matemática con aplicaciones***, México, Prentice Hall.

Canavos, C. George (1998), ***Probabilidad y estadística: aplicaciones y métodos***, México, Mc Graw-Hill.



