



Universidad de Concepción



DEPARTAMENTO DE
KINESOLOGÍA
●● Universidad de Concepción

ESTADÍSTICA DESCRIPTIVA

Klgo. Iván Rodríguez N. MSc, PhD

INTRODUCCIÓN

- La estadística descriptiva busca describir las características biodemográficas de la muestra de estudio.
- Considerar el tipo de dato.
- Es posible utilizar tablas o gráficos.
- La descripción de las variables se realiza en función de:
 - Medida de tendencia central.
 - Medida de dispersión.

TENDENCIA CENTRAL

- Nos indican un valor representativo del grueso de los datos, un valor central.
- Informan sobre cómo son las observaciones prototípicas.
- Los 3 índices de tendencia central más comunes son:
 - **Moda**
 - **Media**
 - **Mediana**

TENDENCIA CENTRAL

•MODA:

- Observación mas frecuente.
- Puede ser usado con todos los tipos de datos.
- No es necesariamente única (bimodal, polimodal)
- Se puede calcular con datos en escala nominal.
- En su cálculo no intervienen todos los elementos.

Eur J Phys Rehabil Med. 2014 Sep 5. [Epub ahead of print]

Effects of a muscular training program on Chronic Obstructive Pulmonary Disease patients with moderate or severe exacerbation antecedents.

López-García A¹, Souto-Camba S, Aparicio-Blanco M, González-Doniz L, Saleta JL, Verea-Hernando H.

⊕ **Author information**

COPD Degree (GOLD)

Degree II 4%

Degree III 56%

Degree IV 40%

LA MODA SE PUEDE EXPRESAR EN VALORES ABSOLUTOS O PORCENTAJES

TENDENCIA CENTRAL

•MEDIANA:

- La Mediana se define como el valor que tiene la propiedad de que el número de observaciones menores que él es igual al número de observaciones mayores que él (Percentil 50).
- Por ejemplo, en la secuencia (ordenada) 3,4,5,6,7,8,9= la mediana será 6.
- En la secuencia (ordenada) 2,3,4,6,7,9
- la mediana será 5 (la media aritmética entre los dos valores centrales)

TENDENCIA CENTRAL

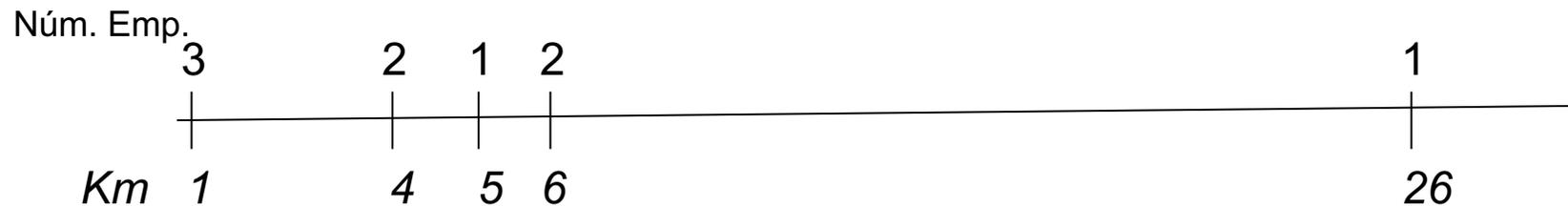
•**MEDIANA:**

- No utiliza todos los elementos.
- Se puede calcular con **datos ordinales**.
- Se ve **menos afectada por datos atípicos** que la media aritmética.
- Minimiza la suma de diferencias en valor absoluto (recordad que la media aritmética minimizaba la suma de diferencias en términos cuadráticos)

TENDENCIA CENTRAL

Ejemplo uso de la mediana

Los 9 empleados de una nueva empresa viven al lado de la ruta 5 en diferentes kilómetros:



Dado que todos viajan en auto, y sabiendo que quieres minimizar el coste en bencina, ¿en qué lugar pondrías la empresa para minimizar tal coste?

TENDENCIA CENTRAL

MEDIA ARITMETICA

Fórmula: $\bar{X} = \frac{\sum X_i}{n}$

- Simplemente se trata de sumar todos los valores y dicha cantidad se divide por el número de valores que tengamos.
- Si tenemos los datos: 4,6,5,3,7 → La media será $(4+6+5+3+7)/5=4$
- Promedios ponderados (Meta-análisis).

Autor	n	r	r_ponderado
Richardson 2012	50	0,7	0,24
Lambert 2000	70	0,5	0,24
Eston 2012	28	0,8	0,15
Total	148		

$$\bar{X} = \frac{0.7 + 0.5 + 0.8}{3} = 0.66$$

r = 0.63

TENDENCIA CENTRAL

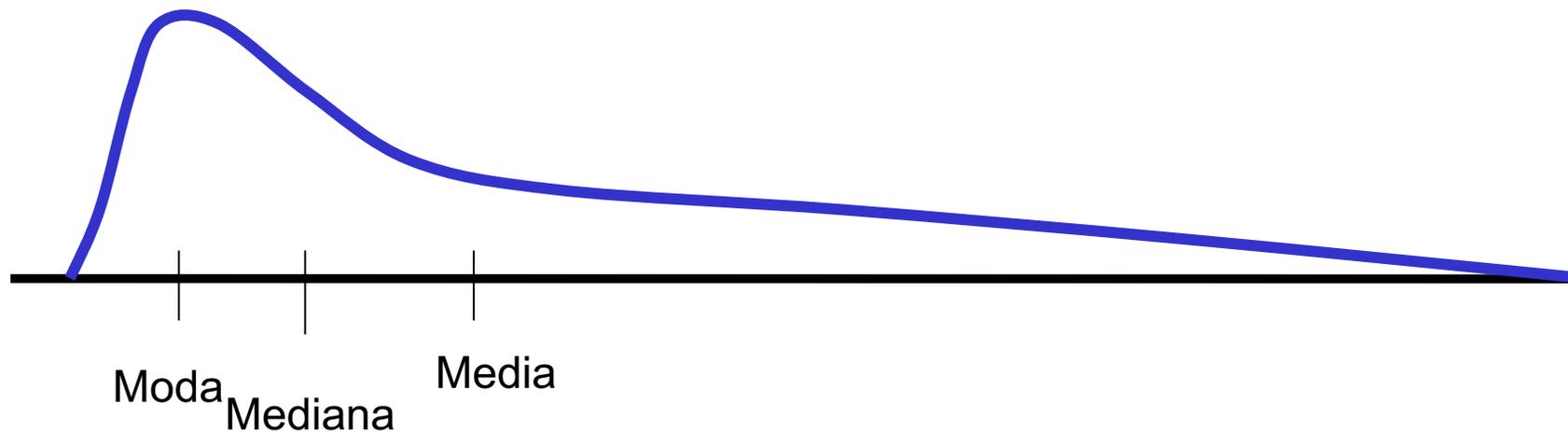
MEDIA ARITMETICA

- La suma de diferencias (de todos los valores) respecto a la media es siempre 0.
- Si sumamos una constante a cada uno de los valores, la nueva media aritmética resultante será la original más la constante.
- Si multiplicamos cada uno de los valores por una constante, la nueva media aritmética será la original por la constante.

RESISTENCIA Y ROBUSTEZ

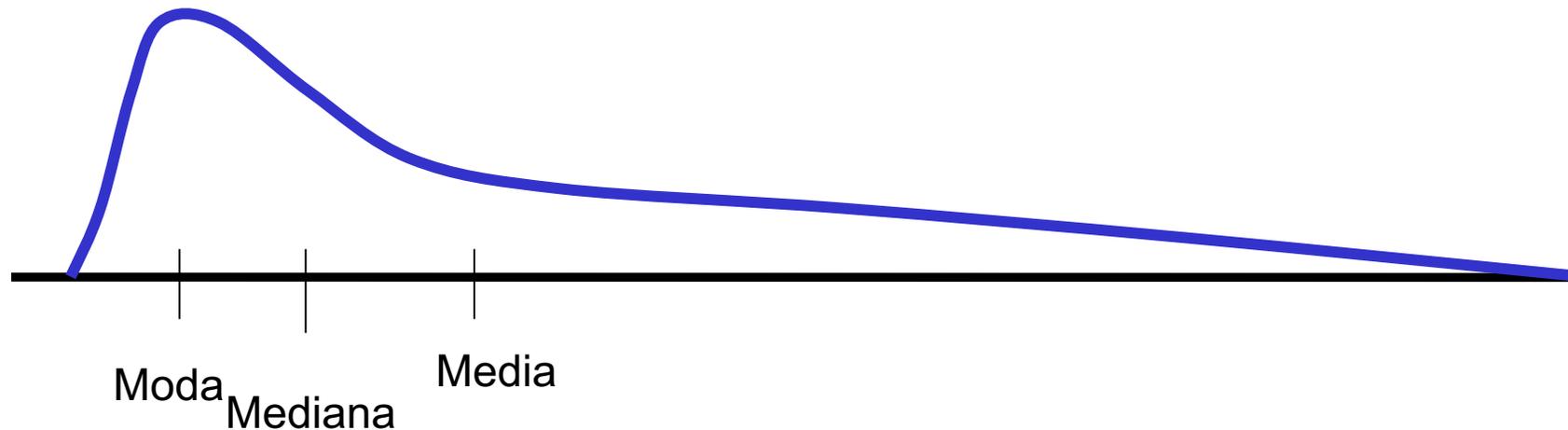
Estadísticos Resistentes: Son aquellos que no se ven influidos por pequeños cambios en los datos.

La mediana es un estadístico altamente resistente.



RESISTENCIA Y ROBUSTEZ

Estimadores robustos: Son aquellos estadísticos (estimadores) que funcionan bien para varios tipos distintos de distribuciones teóricas.



MEDIAS ROBUSTAS DE TENDENCIA CENTRAL

MEDIA RECORTADA

- Media aritmética calculada sobre un subconjunto central de los datos, no considerándose una determinada proporción por cada extremo.
- Por ejemplo, una media recortada al 40% en una secuencia de 10 datos implica no tener en cuenta ni los 4 valores menores ni los 4 valores mayores.
- Observar que la media recortada al 0% es la media aritmética.
- A la media recortada al 25% se la denomina centrimedia.

VARIABILIDAD

MUESTRA A	MUESTRA B
3	13
10	12
20	14
1	11
5	15
50	16
7	10
13,7± 17,1	13± 2,1

VARIABILIDAD

Para el calculo de la variabilidad se debe calcular a que distancia se encuentran las mediciones individuales de la media aritmética.

MUESTRA A	Distancia de la X
3	-10,7
10	-3,7
20	+6,3
1	-12,7
5	-8,7
50	+36,3
7	-6,7
13,7± 17,1	0

VARIABILIDAD

Para el calculo de la variabilidad se debe calcular a que distancia se encuentran las mediciones individuales de la media aritmética.

MUESTRA A	Distancia de la X
3	10,7
10	3,7
20	6,3
1	12,7
5	8,7
50	36,3
7	6,7
13,7± 17,1	DS 17,1

VARIABILIDAD

- La suma de las desviaciones por encima del promedio es igual a la suma de las desviaciones por debajo del promedio.

$$\sum_{i=1}^n (X_i - \bar{X}) = 0$$

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

CALCULANDO LA VARIANZA Y DESVIACIÓN ESTÁNDAR

x	x-X	(x-X) ²
3	-10,7	114,49
10	-3,7	13,69
20	6,3	39,69
1	-12,7	161,29
5	-8,7	75,69
50	36,3	1317,69
7	-6,7	44,89
X=13,7	Σ=0	Σ=1767,43

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

$$s^2 = \frac{1767,43}{6}$$

CALCULANDO LA VARIANZA Y DESVIACIÓN ESTÁNDAR

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

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

- La desviación típica viene dada en las mismas unidades de medida que los datos originales (en la varianza las unidades están al cuadrado).
- Por eso, en estadística descriptiva se suele dar la media acompañada de la desviación típica.

OTRAS MEDIDAS DE DISPERSIÓN

RANGO

- Diferencia entre el valor máximo y mínimo de una muestra.
- Es altamente influenciado por magnitud de valores extremos.

FC
120
125
138
123
151
135
132
155
154
137
135
132
139
135
101
129
155
99
125
110
136
135

MEDIDAS DE DISPERSIÓN

RANGO

	N	Mín	Máx	Media	DS
FC	22	99,0	155,0	131,8	15,3
N válido (según lista)	22				

FC
120
125
138
123
151
135
132
155
154
137
135
132
139
135
101
129
155
99
125
110
136
135

MEDIDAS DE DISPERSIÓN

RANGO

	N	Mín	Máx	Media	DS
FC	22	40,0	210,0	130,7	30,10
N válido (según lista)	22				

FC
120
125
138
123
151
40
132
155
154
137
135
132
210
135
101
129
155
99
125
110
136
135

OTRAS MEDIDAS DE DISPERSIÓN

RANGO INTERCUARTILICO

- Rango entre el percentil 25 y 75.

	Percentiles						
	5	10	25	50	75	90	95
FC	99,3	103,7	124,5	135,0	138,2	154,7	155,0

FC
99
101
110
120
123
125
125
129
132
132
135
135
135
135
136
137
138
139
151
154
155
155

RANGO INTERCUARTILICO

- No es influenciado por valores extremos.

	Percentiles						
	5	10	25	50	75	90	95
FC1	99,3	103,7	124,5	135,0	138,2	154,7	155,0
FC2	48,85	99,0	122,2	133,5	141,2	155,0	201,7

FC 1	FC 2
99	40
101	99
110	101
120	110
123	120
125	123
125	125
129	125
132	129
132	132
135	132
135	135
135	135
135	135
135	135
136	136
137	137
138	138
139	151
151	154
154	155
155	155
155	210

RANGO INTERCUARTILICO

- No es influenciado por valores extremos.

		Estadístico
FC1	Media	131,8636
	Media recortada al 5%	132,3939
	Mediana	135,0000
	Varianza	234,790
	Desv. típ.	15,32286
	Mínimo	99,00
	Máximo	155,00
	Rango	56,00
FC2	Media	130,7727
	Media recortada al 5%	131,3939
	Mediana	133,5000
	Varianza	923,232
	Desv. típ.	30,38473
	Mínimo	40,00
	Máximo	210,00
	Rango	170,00

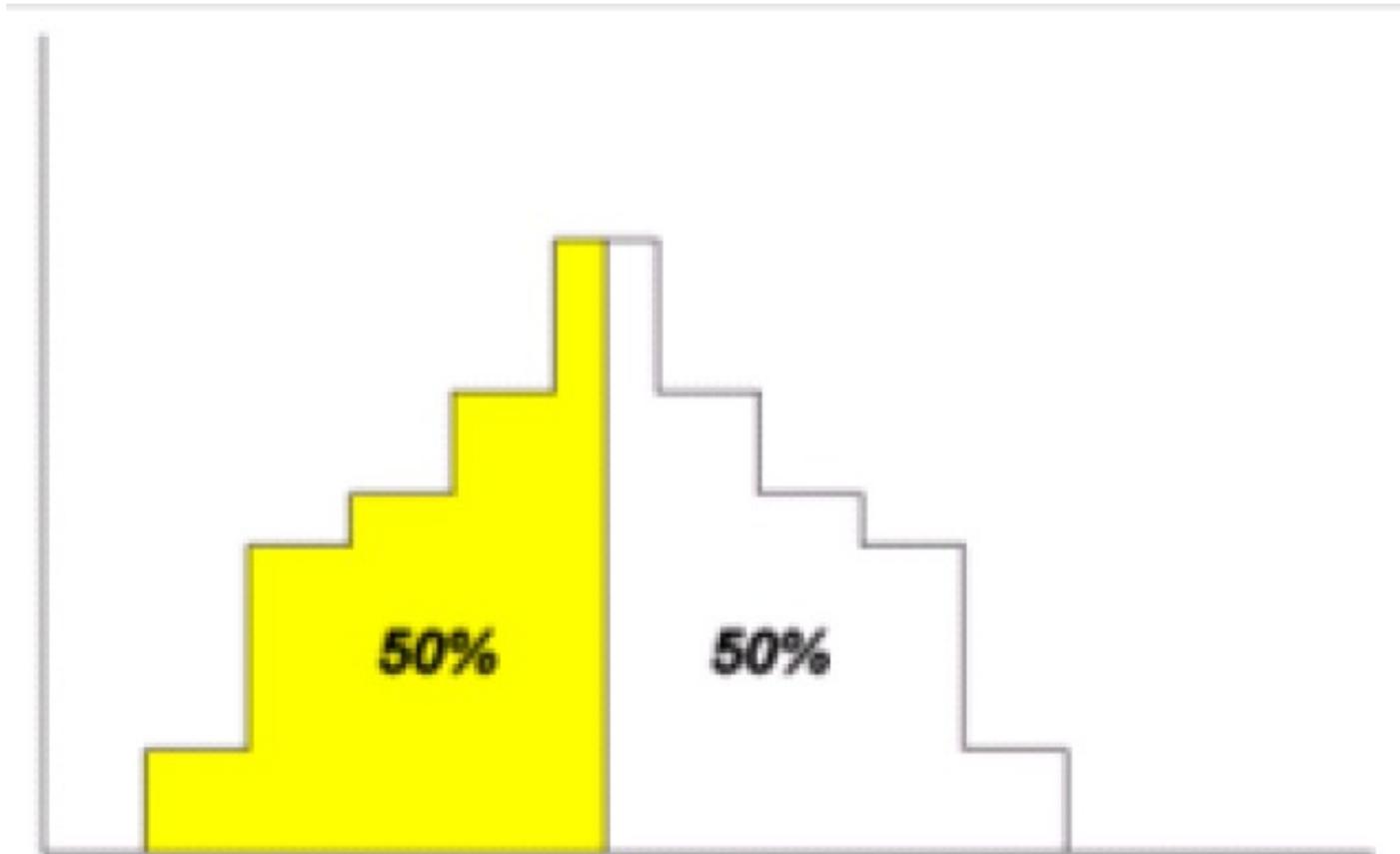
FC 1	FC 2
99	40
101	99
110	101
120	110
123	120
125	123
125	125
129	125
132	129
132	132
135	132
135	135
135	135
135	135
136	136
137	137
138	138
139	151
151	154
154	155
155	155
155	210

CARACTERÍSTICA DE UNA DISTRIBUCIÓN (NORMAL)

- ASIMETRÍA

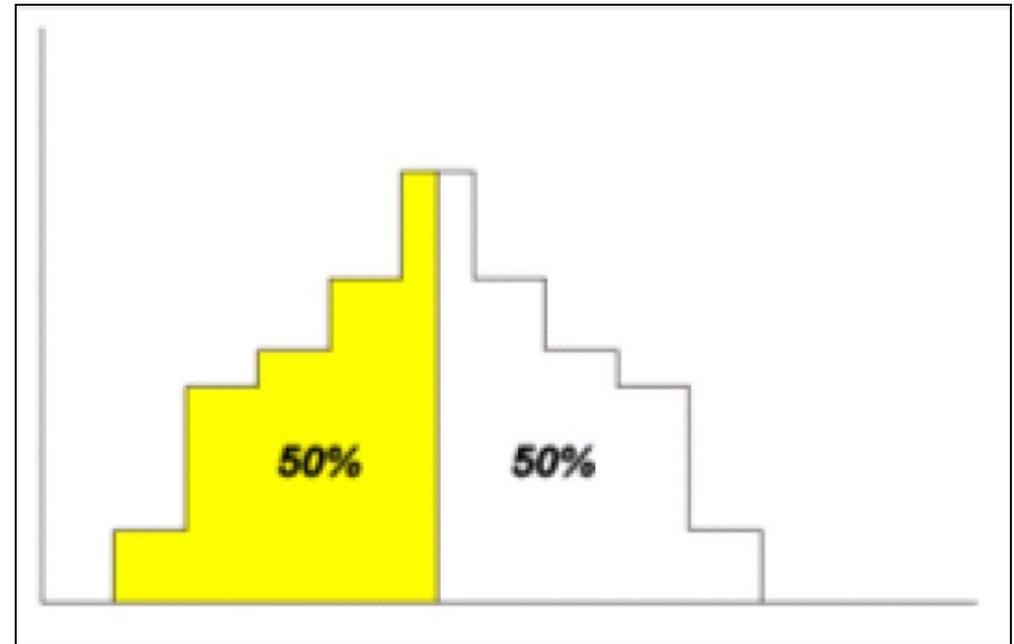
- Grado en que valores de la variable, equidistantes al centro de la distribución, poseen frecuencias similares.
- Coeficiente de asimetría en una distribución normal es=0.
- El eje de simetría es la media (mediana).

CARACTERÍSTICA DE UNA DISTRIBUCIÓN (NORMAL SIMÉTRICA)

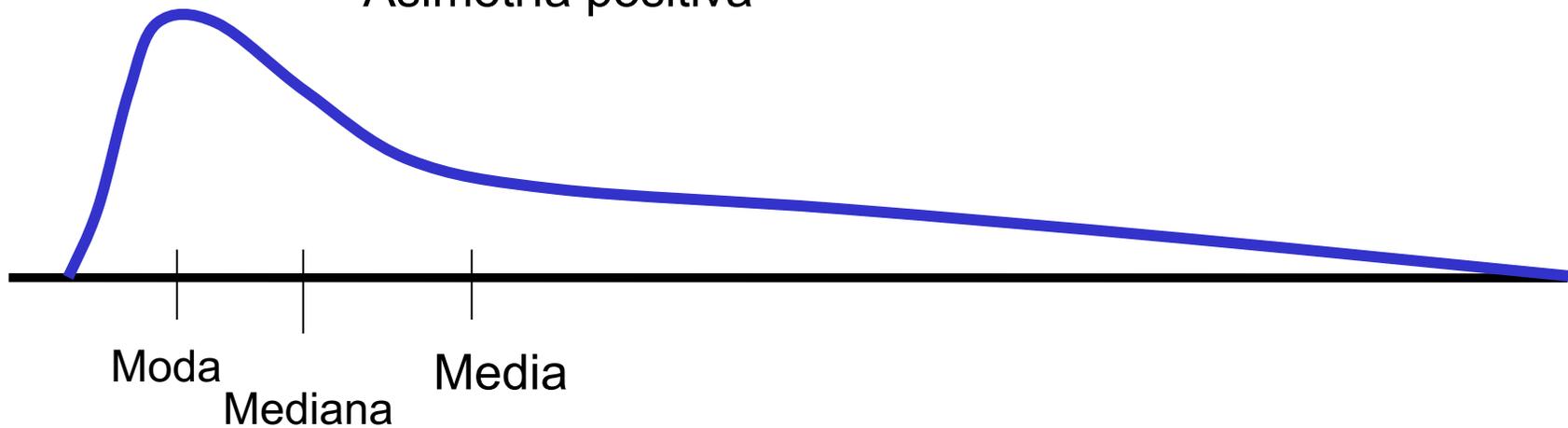


CARACTERÍSTICA DE UNA DISTRIBUCIÓN (NORMAL)

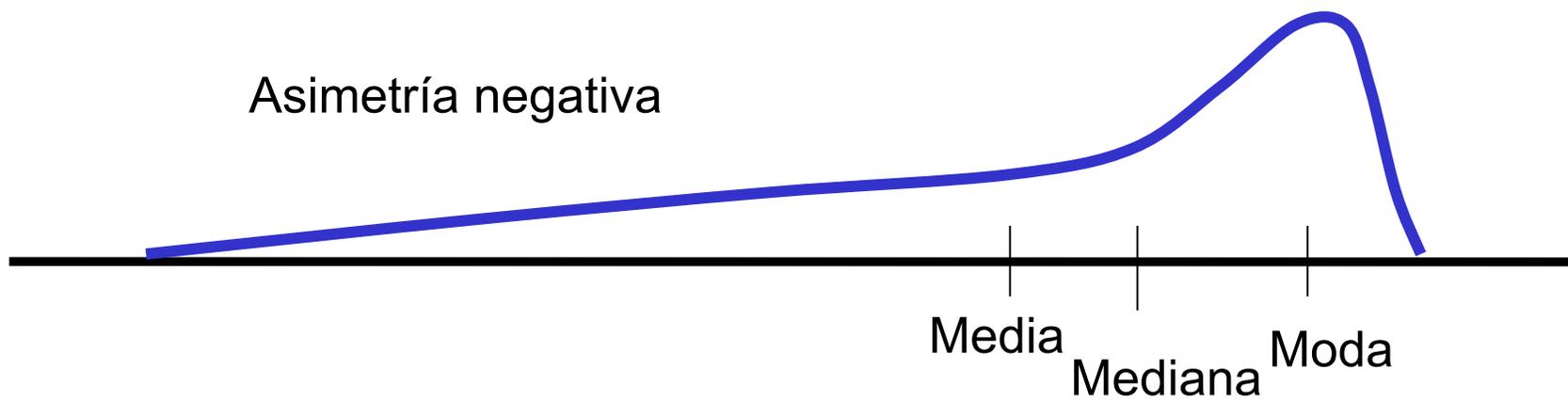
- A través de la observación del histograma de frecuencias es posible determinar si la distribución es simétrica o no.
- Cuando la distribución de los datos es simétrica, la media, la mediana y la moda coinciden. (Y la distribución tiene la misma forma a la izquierda y la derecha del centro).



Asimetría positiva



Asimetría negativa



INDICES DE ASIMETRÍA

ÍNDICE DE ASIMETRÍA DE PEARSON

Está basado en la relación entre la media y la moda en distribuciones simétricas y asimétricas:

$$A_s = \frac{\bar{X} - Mo}{S_x}$$

Si la distribución es simétrica A_s será 0

Si la distribución es asimétrica positiva, A_s será mayor que 0

Si la distribución es asimétrica negativa, A_s será menor que 0

INDICES DE ASIMETRÍA

ÍNDICE DE ASIMETRÍA DE FISHER

Está basado en la diferencia de los datos sobre la media, como la varianza, si bien esta vez se elevan los coeficientes al cubo

$$A_s = \frac{\sum_{i=1}^n (X_i - \bar{X})^3 / n}{s_x^3}$$

Si la distribución es simétrica A_s será 0

Si la distribución es asimétrica positiva, A_s será mayor que 0

Si la distribución es asimétrica negativa, A_s será menor que 0

Desventaja: Muy influida por puntuaciones atípicas.

CURTOSIS O APUNTAMIENTO

- Hace referencia al apuntamiento de la distribución en relación a un estándar, que es la distribución normal (mesocurtica).
 - Distribución leptocurtica.
 - Distribución platicurtica

INDICE DE CURTOSIS

Para una distribución normal (mesocúrtica) sabemos que:

$$C_r = \frac{\sum_{i=1}^n (X_i - \bar{X})^4 / n}{s_x^4} - 3$$

Si la distribución es normal (mesocúrtica), el índice vale 0

Si la distribución es leptocúrtica, el índice es superior a 0

Si la distribución es platicúrtica, el índice es inferior a 0

REPORTANDO RESULTADOS DESCRIPTIVOS

- TABLAS
- GRÁFICOS

TABLAS

Tabla 1. Características biodemográficas de la muestra.

Características generales	ENM (n=18)	DPC (n=11)
Edad (años)	12 (5-17)	13 (5-16)
Sexo (Hombre/Mujer)	11/6	6/5
Peso (kg)	41 (23-88)	38 (18-71)
Talla (m)	1,45 (1,10-1,60)	1,42 (1,03-1,61)
Camina/no camina	8/10	11/0
Duración EMR (semanas)	39 (24-89)	60 (46-90)
Función pulmonar		
VEF ₁ (L)	1,67 (0,49 – 2,95)	1,35 (0,55 – 2,67)
% VEF ₁	79,0 (30 – 116)	54,0 (30 – 99)
CVF (L)	1,85 (0,58 – 3,28)	1,92 (0,65 – 3,43)
%CVF	78,5 (11 – 114)	81,0 (49 – 112)
VEF ₁ /CVF	88,5 (71 – 100)	64,0 (39 – 85)
FEF ₂₅₋₇₅ (L/seg)	1,83 (0,5 – 5,9)	0,57 (0,26 – 2,29)
%FEF ₂₅₋₇₅	66,0 (30 – 142)	31,0 (11 – 72)

ENM: Enfermedad neuromuscular; DPC: Daño pulmonar crónico; VEF₁: Volumen espiratorio forzado en el primer segundo; CVF: Capacidad vital forzada; FEF₂₅₋₇₅: Flujo espiratorio forzado entre el 25 y 75 de la CVF. Resultados se expresan en mediana y rango.

TABLAS DE FRECUENCIA

TABLE 2.10

Absolute frequencies of birth weights for 3,751,275 infants born in the United States, 1986

Birth Weight (grams)	Number of Infants
0-499	4843
500-999	17,487
1000-1499	23,139
1500-1999	49,112
2000-2499	160,919
2500-2999	597,738
3000-3499	1,376,008
3500-3999	1,106,634
4000-4499	344,390
4500-4999	62,769
5000-5500	8236
Total	3,751,275

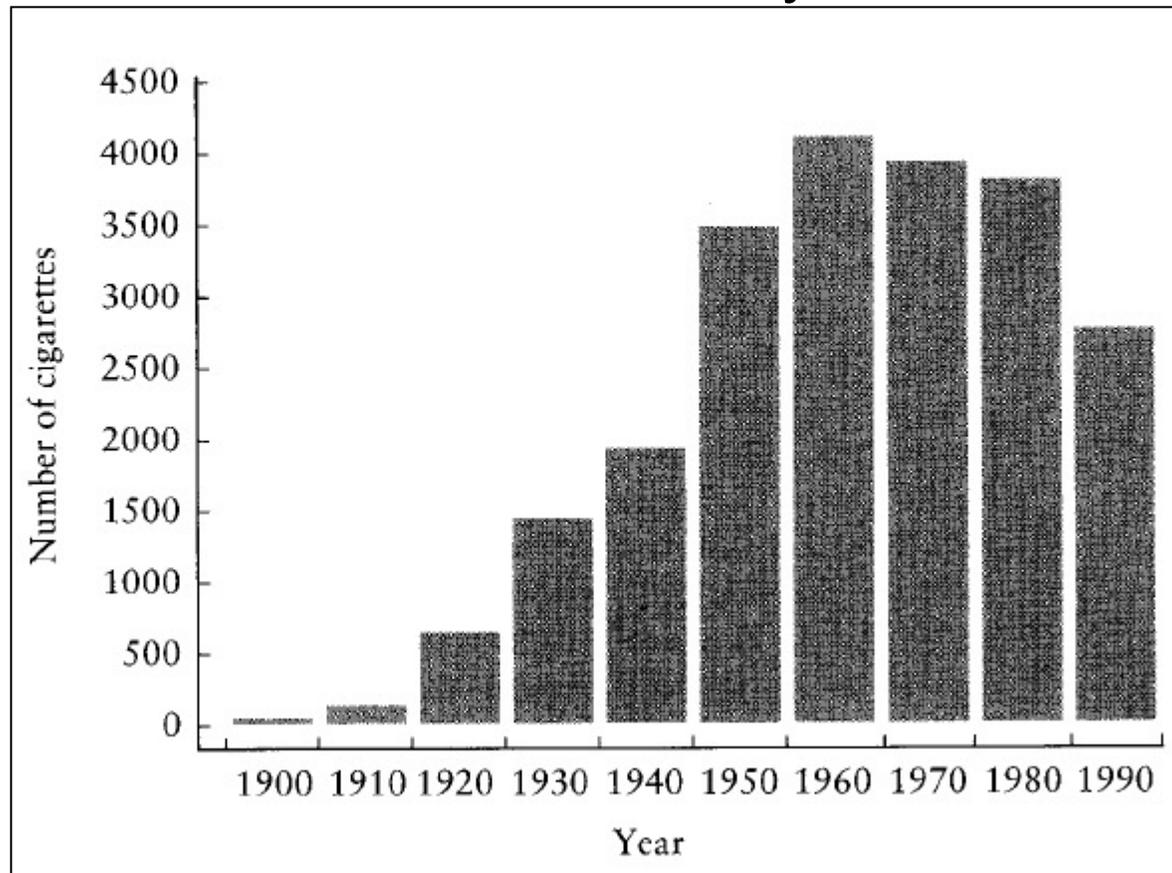
TABLE 2.11

Relative frequencies of birth weights for 3,751,275 infants born in the United States, 1986

Birth Weight (grams)	Relative Frequency (%)
0-499	0.1
500-999	0.5
1000-1499	0.6
1500-1999	1.3
2000-2499	4.3
2500-2999	15.9
3000-3499	36.7
3500-3999	29.5
4000-4499	9.2
4500-4999	1.7
5000-5500	0.2
Total	100.0

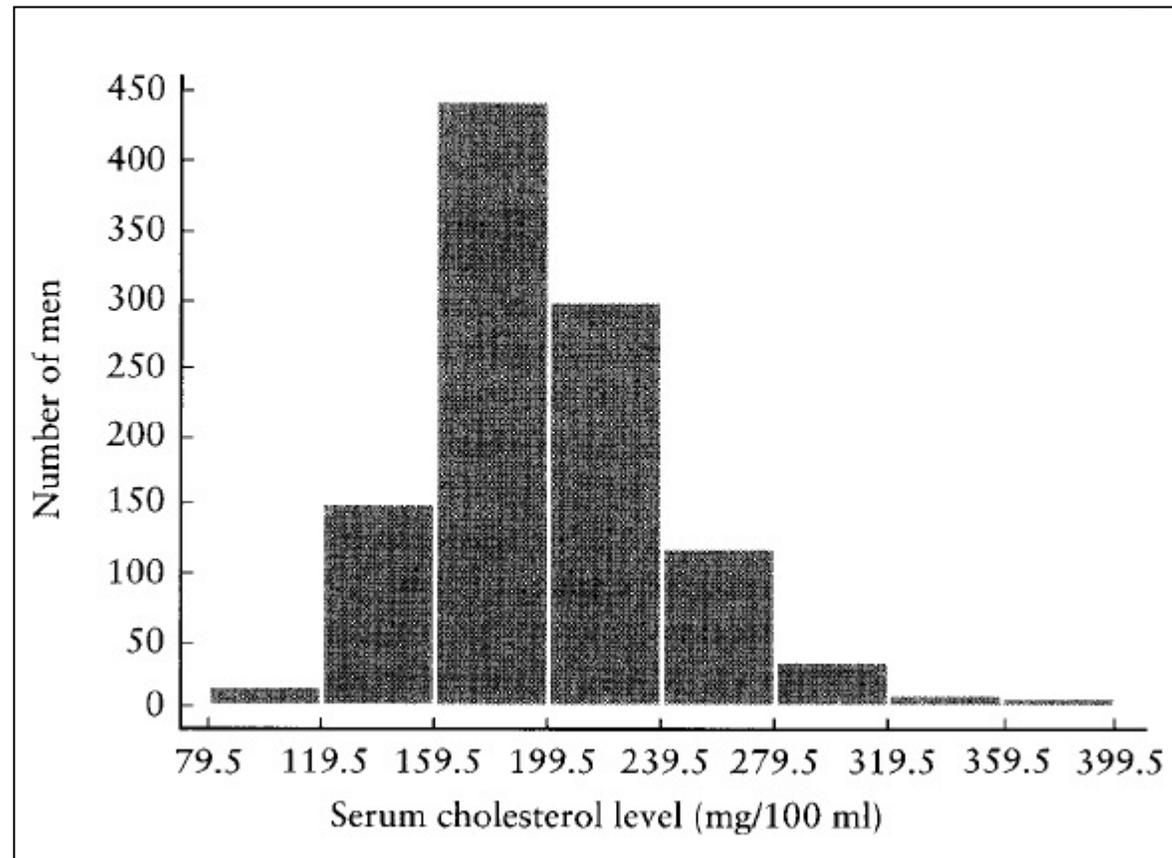
GRAFICO DE BARRA

- Ampliamente utilizado para graficar distribución de frecuencias de datos nominales y ordinales.

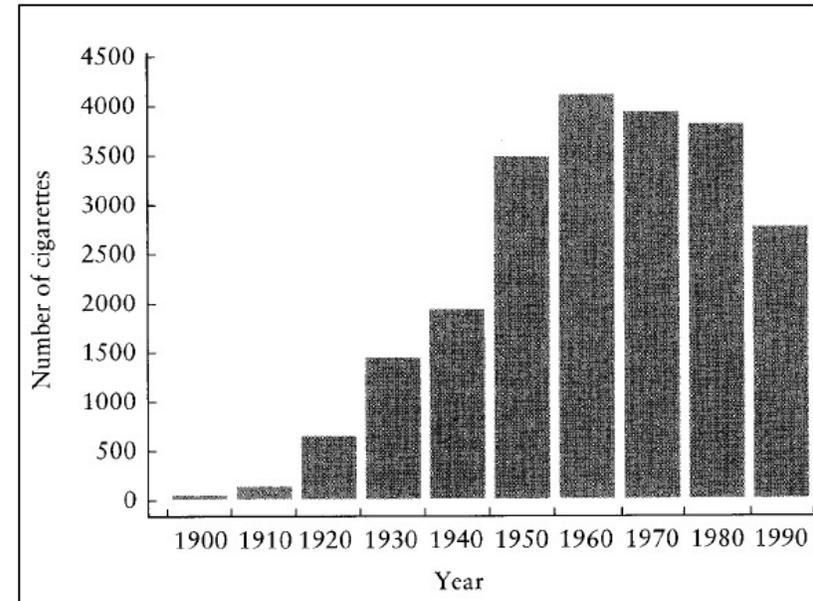
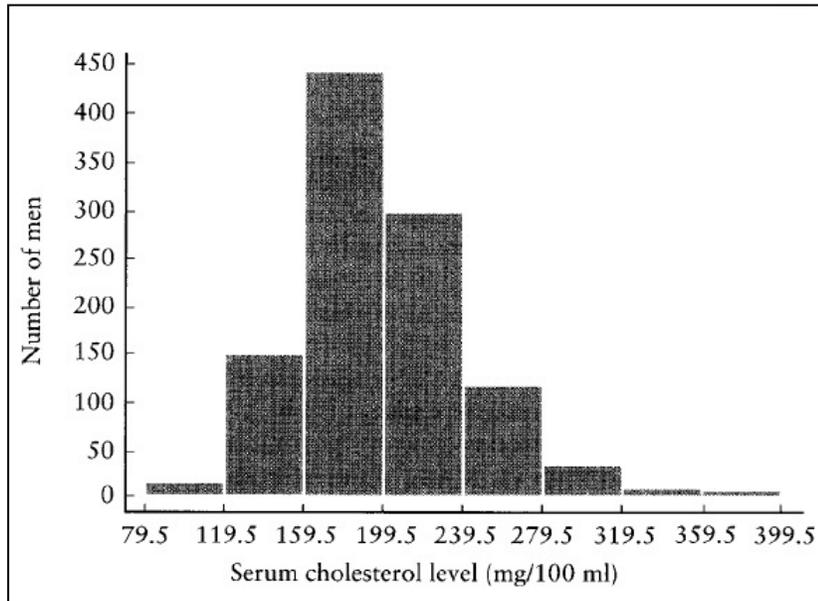


HISTOGRAMA

- Utilizado para graficar distribución de frecuencias de datos continuos y discretos.

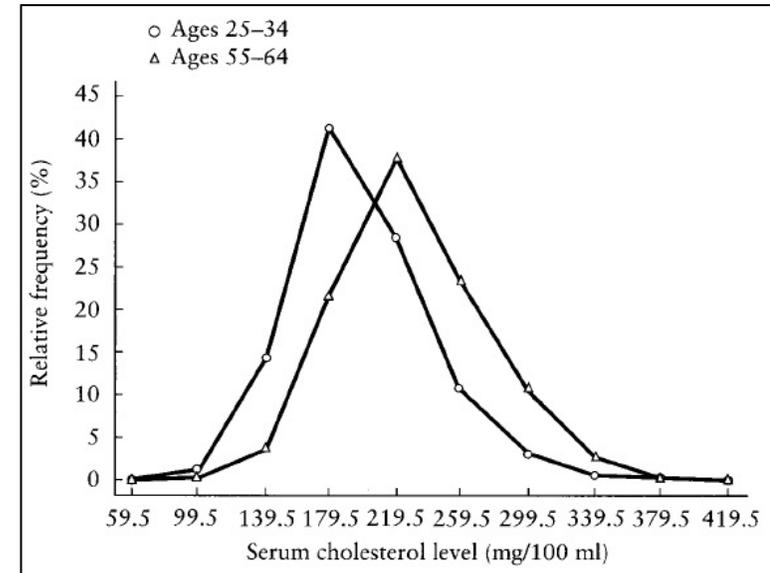
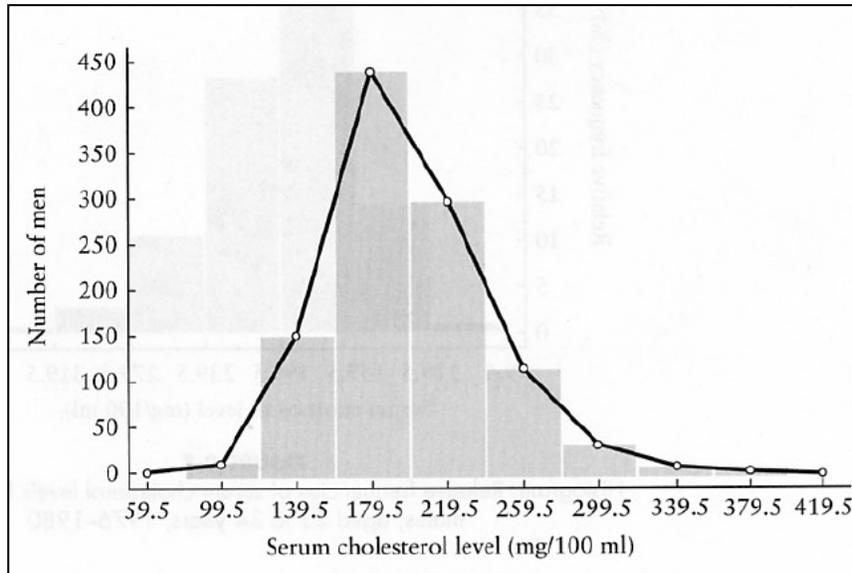


HISTOGRAMA/GRAFICO DE BARRAS



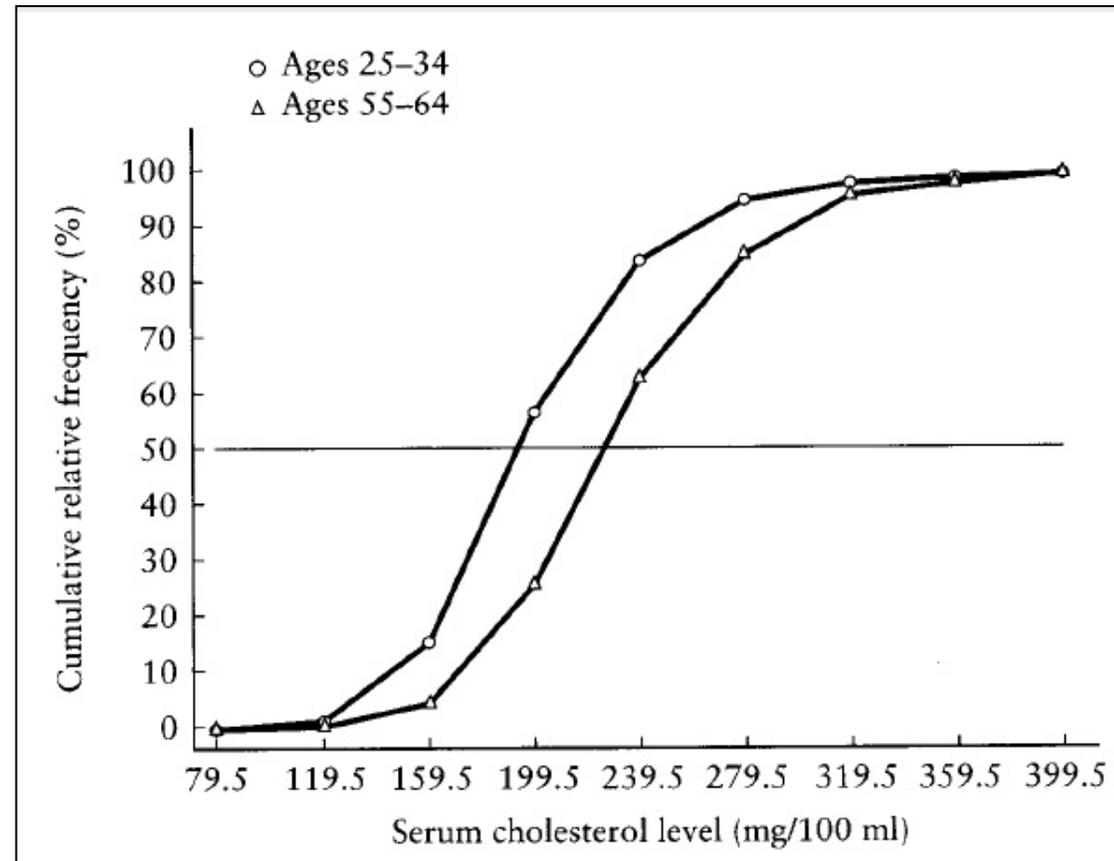
- En histograma eje X representa los límites de los intervalos de una variable continua.
- El área total ocupada por las barras representa 1.

POLIGONO DE FRECUENCIAS



- Útil en la comparación de distribuciones.

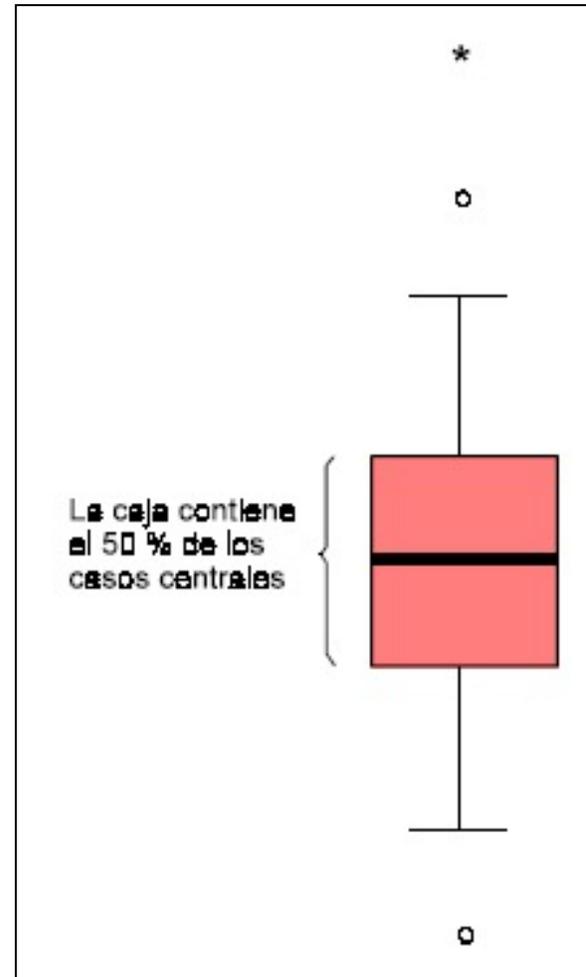
POLIGONO DE FRECUENCIAS ACUMULADAS



- Útil en la comparación de distribuciones.

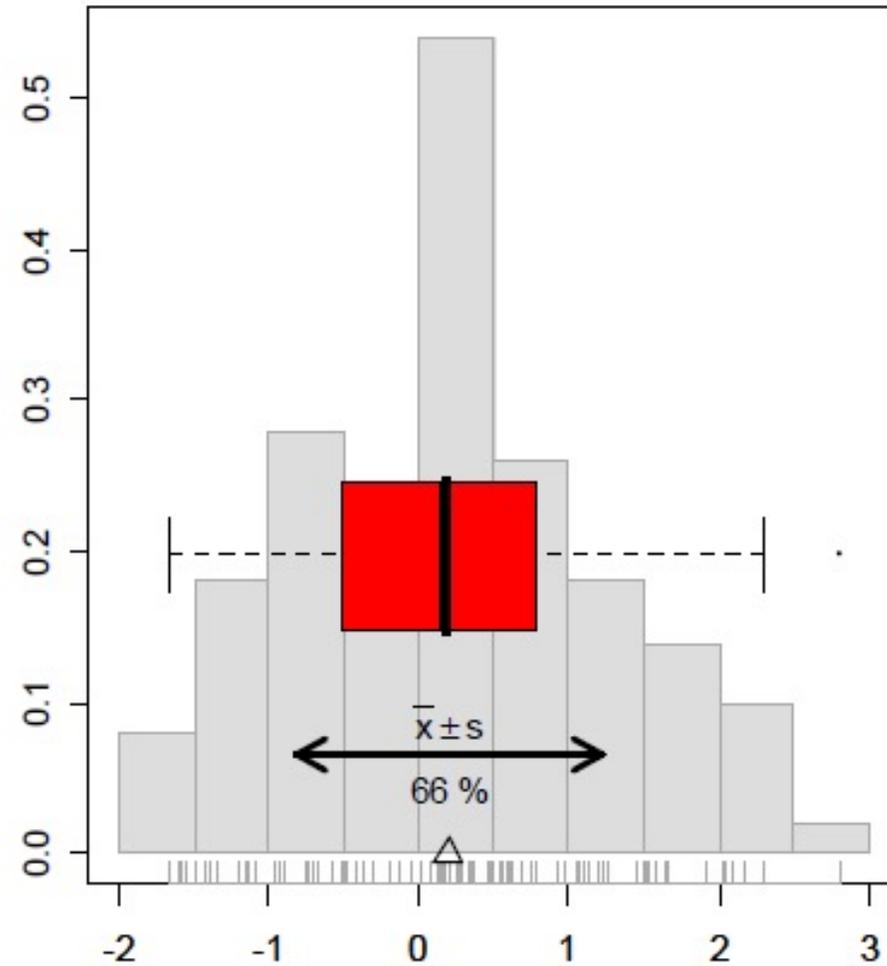
BOX PLOTS

- Muestran resumen de los datos.
- Caja se extiende entre el percentil 25-75.
- Línea representa percentil 50 (mediana).
- Líneas extremas representan valores adyacentes.
- Valor adyacente: punto que queda 1,5 veces el rango intercuartil por encima o por debajo del percentil 75 y 25 respectivamente.

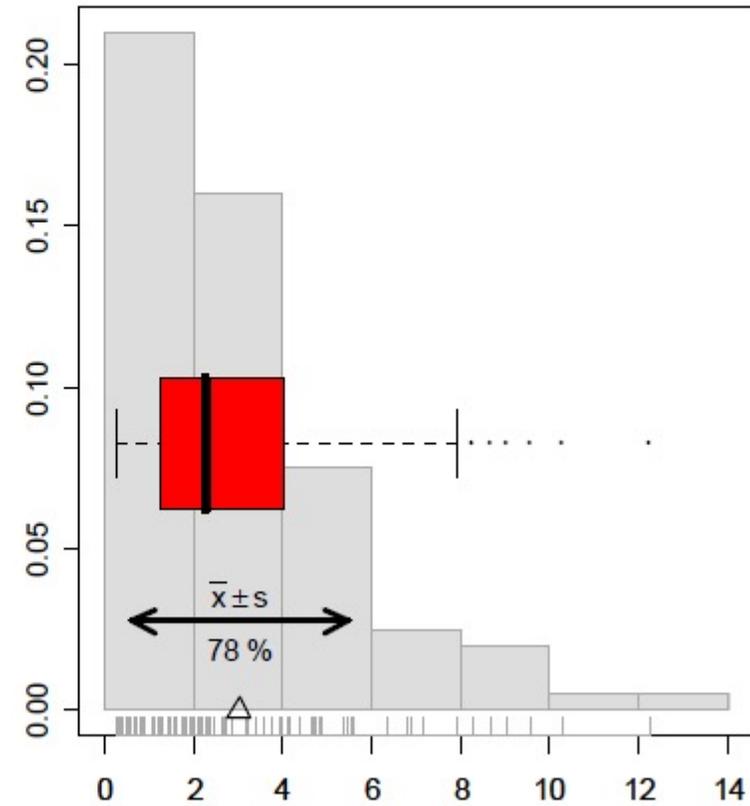
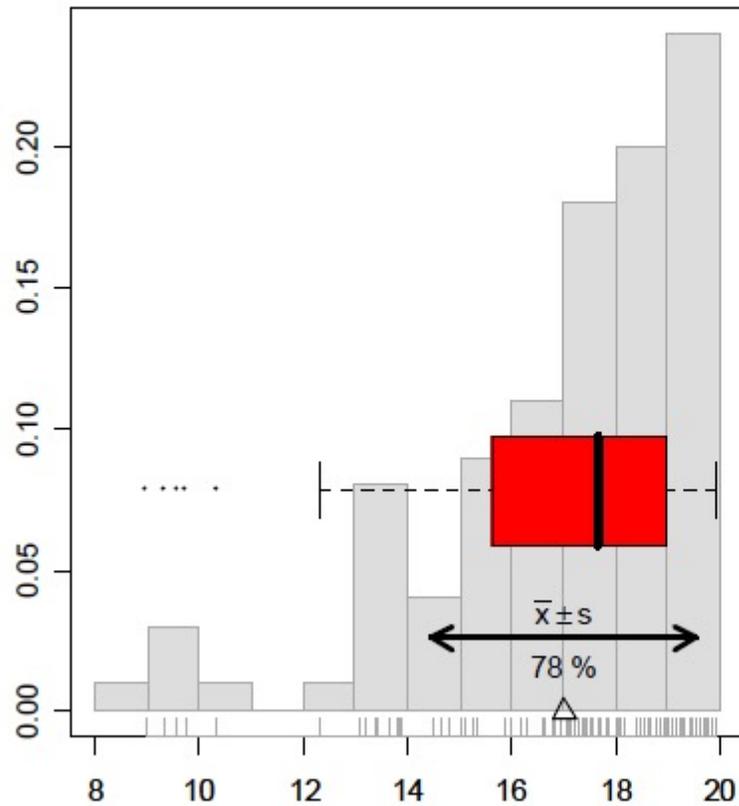


BOX PLOTS Y DISTRIBUCION DE FRECUENCIA

DISTRIBUCION NORMAL SIMETRICA.

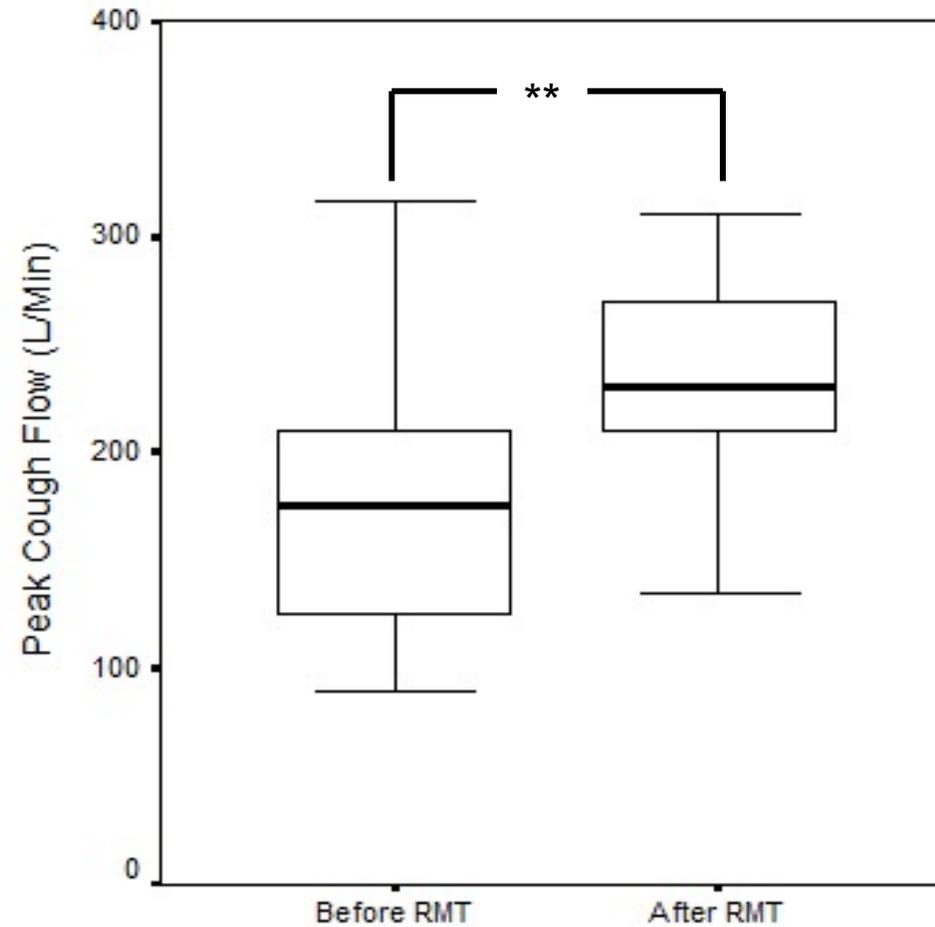


BOX PLOTS Y DISTRIBUCION DE FRECUENCIA



Effects of home respiratory muscle training in children with chronic respiratory disease.

Iván Rodríguez N^{1,2,4}, Daniel Zenteno A^{1,3}, Carlos Manterola D^{4,5}.



Figures 2. Effect of respiratory muscle training (RMT) on peak cough flow in children with neuromuscular disease (n= 18). $p < 0.05^{**}$:



Universidad de Concepción



DEPARTAMENTO DE
KINESOLOGÍA
●● Universidad de Concepción

ESTADÍSTICA DESCRIPTIVA

Klgo. Iván Rodríguez N. MSc, PhD