



FUNDACIÓN UNIVERSITARIA  
DE CIENCIAS DE LA SALUD  
**FUCS**

# Fundamentos de Bioestadística

Camilo Yate Támara

[camilo\\_yate@yahoo.fr](mailto:camilo_yate@yahoo.fr)

Tel: 3017662953

# Introducción

Los avances tecnológicos en las ciencias de la computación han dado vuelco a la forma tradicional de hacer estadística. La tecnología ha permitido aplicar lo que se conoce como “estadística moderna” donde la importancia **no radica en la memorización** y el cálculo aritmético manual de las formulas sino en enfocarse en obtener resultados para tener sentido practico de los datos mediante análisis critico.

En la siguiente sección se definirán varios conceptos que van atados a definiciones y notación matemática. Sin embargo, no es la intención de este curso la memorización ni la realización exhaustiva de cálculos matemáticos. Solo se presentarán fórmulas matemáticas para dar la formalidad de la definición a cada concepto.

Los cálculos de los conceptos mostrados en esta sección serán desarrollados de manera práctica en un taller en clase en softwares diseñados para tal fin.

# Tablas de Frecuencias



# Tablas de Frecuencias

Cuando se trabajan con conjuntos de datos extensos es útil organizar y resumir la información mediante una tabla que liste todos los valores posibles junto con la frecuencia de ocurrencia de los mismos. Para esto se definen cuatro tipos de frecuencia

**Frecuencia Absoluta:** El conteo en cada valor posible de la variable ( $f$ )

**Frecuencia Relativa:** Es el porcentaje de participación de cada categoría en el total ( $f_{\Delta} = \frac{f}{n}$ )

**Frecuencia Absoluta Acumulada:** Es el resultado de ir sumando las frecuencias absolutas de los valores de una variable ( $F$ )

**Frecuencia Relativa Acumulada:** Es el cociente entre la frecuencia acumulada y el número total de datos ( $F_{\Delta} = \frac{F}{n}$ )

Las forma de construir una tabla de frecuencia se construyen según el tipo de variable (cualitativa o cuantitativa) a continuación se mostrará la forma de construcción de cada una de ellas

# Tablas de Frecuencias

## Variables Cuantitativas

Suponga que se cuenta con los datos de la presión arterial sistólica de 25 pacientes.

Individuo	Presion Arterial Sistolica (mm Hg)
1	111
2	136
3	133
4	105
5	157
6	80
7	154
8	96
9	114
10	126
11	106
12	146
13	128
14	112
15	137
16	91
17	94
18	102
19	133
20	132
21	119
22	91
23	121
24	137
25	125

Datos Simulados

1

Elegir el numero de grupos (clases:  $k$ ). En nuestro ejemplo  $k = 5$

2

Calcular el ancho de clase con la formula

$$w = \frac{\text{dato maximo} - \text{dato minimo}}{k} = \frac{157 - 80}{5} = 15.4 \approx 16$$

3

Construir los intervalos a partir del dato menor y adicionando el valor de  $w$  sucesivamente

Clase	Presion Arterial Sistolica (mm Hg)	
1	80	96
2	97	113
3	114	130
4	131	147
5	148	164
Total		

# Tablas de Frecuencias

## Variables Cuantitativas

Suponga que se cuenta con los datos de la presión arterial sistólica de 25 pacientes.

Individuo	Presion Arterial Sistolica (mm Hg)
1	111
2	136
3	133
4	105
5	157
6	80
7	154
8	96
9	114
10	126
11	106
12	146
13	128
14	112
15	137
16	91
17	94
18	102
19	133
20	132
21	119
22	91
23	121
24	137
25	125

Datos Simulados

3

Calcular la frecuencia absoluta contando el numero de casos en cada intervalo

Clase	Presion Arterial Sistolica (mm Hg)		Frecuencia Absoluta
1	80	96	5
2	97	113	5
3	114	130	6
4	131	147	7
5	148	164	2
Total			25

4

Calcular la frecuencia absoluta acumulada a partir de la frecuencia absoluta

Clase	Presion Arterial Sistolica (mm Hg)		Frecuencia Absoluta	Frecuencia Absoluta Acumulativa
1	80	96	5	5
2	97	113	5	10
3	114	130	6	16
4	131	147	7	23
5	148	164	2	25
Total			25	

# Tablas de Frecuencias

## Variables Cuantitativas

Suponga que se cuenta con los datos de la presión arterial sistólica de 25 pacientes.

Individuo	Presion Arterial Sistolica (mm Hg)
1	111
2	136
3	133
4	105
5	157
6	80
7	154
8	96
9	114
10	126
11	106
12	146
13	128
14	112
15	137
16	91
17	94
18	102
19	133
20	132
21	119
22	91
23	121
24	137
25	125

Datos Simulados

5

Calcular la frecuencia relativa dividiendo la frecuencia absoluta en el numero de datos

Clase	Presion Arterial Sistolica (mm Hg)		Frecuencia Absoluta	Frecuencia Absoluta Acumulativa	Frecuencia Relativa
1	80	96	5	5	20%
2	97	113	5	10	20%
3	114	130	6	16	24%
4	131	147	7	23	28%
5	148	164	2	25	8%
Total			25		100%

6

Calcular la frecuencia relativa acumulada a partir de la frecuencia relativa

Clase	Presion Arterial Sistolica (mm Hg)		Frecuencia Absoluta	Frecuencia Absoluta Acumulativa	Frecuencia Relativa	Frecuencia Relativa Acumulativa
1	80	96	5	5	20%	20%
2	97	113	5	10	20%	40%
3	114	130	6	16	24%	64%
4	131	147	7	23	28%	92%
5	148	164	2	25	8%	100%
Total			25		100%	

# Tablas de Frecuencias

## Variables Cualitativas

Suponga que se cuenta con una encuesta con el grupo sanguíneo de una población de 15 individuos.

Individuo	Grupo Sanguineo
1	AB
2	O
3	AB
4	AB
5	B
6	AB
7	O
8	A
9	B
10	AB
11	O
12	B
13	B
14	A
15	AB

Datos Simulados

1

	Frecuencia Absoluta
A	2
B	4
O	3
AB	6
<b>Total</b>	<b>15</b>

2

	Frecuencia Absoluta	Frecuencia Absoluta Acumulativa
A	2	2
B	4	6
O	3	9
AB	6	15
<b>Total</b>	<b>15</b>	

3

	Frecuencia Absoluta	Frecuencia Absoluta Acumulativa	Frecuencia Relativa
A	2	2	13.33%
B	4	6	26.67%
O	3	9	20.00%
AB	6	15	40.00%
<b>Total</b>	<b>15</b>		<b>100.00%</b>

4

	Frecuencia Absoluta	Frecuencia Absoluta Acumulativa	Frecuencia Relativa	Frecuencia Relativa Acumulativa
A	2	2	13.33%	13%
B	4	6	26.67%	40%
O	3	9	20.00%	60%
AB	6	15	40.00%	100%
<b>Total</b>	<b>15</b>		<b>100.00%</b>	

$$f_{\Delta} = \frac{3}{15} = 0.2$$



# Medidas de Tendencia Central



# Medidas de Tendencia Central

## Media Aritmética

Es el promedio de los datos, se obtiene calculando el cociente entre la suma de todos los datos y el número de observaciones

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

## Ejemplo

Se tienen mediciones de la calidad de aire en diferentes zonas de Bogotá medida en ppm

5.40 1.10 0.42 0.73 0.48 1.10

$$\bar{x} = \frac{\Sigma x}{n} = \frac{5.40 + 1.10 + 0.42 + 0.73 + 0.48 + 1.10}{6} = \frac{9.23}{6} = 1.538$$

Existen otras medias de utilidad como **Media ponderada** y **Media Armónica** estas serán definidas en el taller práctico



# Medidas de Tendencia Central

## Mediana

Es el centro de los datos una vez que estos están ordenados. Para calcularla se siguen los siguientes pasos:

1. Ordenar los dato de manera ascendente
2. Si el número de observaciones es impar la mediana es el dato de la mitad
3. Si el número de observaciones es par la mediana es el promedio de los dos datos de la mitad

## Ejemplo

Se tienen mediciones de la calidad de aire en diferentes zonas de Bogotá medida en ppm

5.40 1.10 0.42 0.73 0.48 1.10

Se ordenan los datos

0.42 0.48 0.73 1.10 1.10 5.40

Como el número de observaciones es par

$$\tilde{x} = \frac{0.73 + 1.10}{2} = \frac{1.83}{2} = 0.915$$



# Medidas de Tendencia Central

## Moda

Es el dato de mayor frecuencia, es decir el dato que mas se repite en la muestra

Cuando dos valores comparten la máxima ocurrencia en la muestra, se dice que esta es **bimodal**.

Cuando varios valores comparten la máxima ocurrencia en la muestra, se dice que esta es **multimodal**.

Cuando ningún valor se repite, se dice que no hay moda

## Ejemplo

Suponga las siguientes muestras

a.	5.40	1.10	0.42	0.73	0.48	1.10			
b.	27	27	27	55	55	55	88	88	99
c.	1	2	3	6	7	8	9	10	

# Medidas de Localización



FUNDACIÓN UNIVERSITARIA  
DE CIENCIAS DE LA SALUD  
**FUCS**

# Medidas de Localización

## Percentiles

Indica, una vez ordenados los datos de menor a mayor, el valor de la variable por debajo del cual se encuentra un porcentaje dado de observaciones en un grupo.

## Ejemplo

Se cuentan con los datos del IMC de diez individuos seleccionados en una muestra

23.77    18.77    21.08    18.60    19.16    29.68    22.01    23.75    27.15    21.37

Posición	1	2	3	4	5	6	7	8	9	10
x	18.60	18.77	19.16	21.08	21.37	22.01	23.75	23.77	27.15	29.68

P<sub>10</sub>

P<sub>50</sub>

P<sub>90</sub>

P <sub>10</sub>	18.61
P <sub>50</sub>	21.69
P <sub>90</sub>	29.43

Nótese que el Percentil 50 es equivalente a la mediana.  
Los percentiles más usados son el 10 (o primer decil) , 25 (primer cuartil), 50 (mediana o segundo cuartil), 75 (tercer cuartil) y 90 (novenio decil)

# Medidas de Dispersión



# Medidas de Dispersión

## Rango

Es la diferencia entre el dato mayor y el dato menor

## Ejemplo

5.40 1.10 0.42 0.73 0.48 1.10

$$\text{rango} = 5.4 - 0.42 = 4.98$$

## Rango Inter cuartil

Es la diferencia entre el percentil 75 y el percentil 25

5.40 1.10 0.42 0.73 0.48 1.10

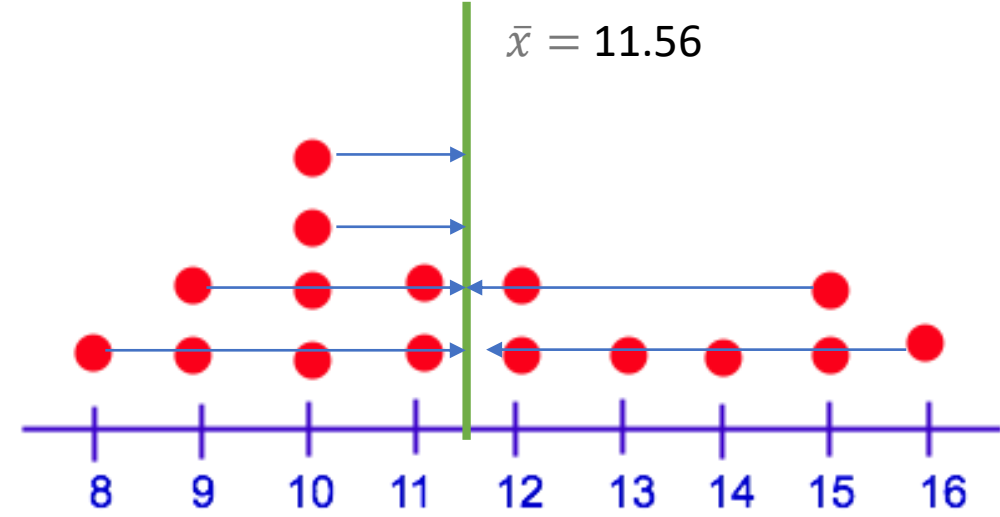
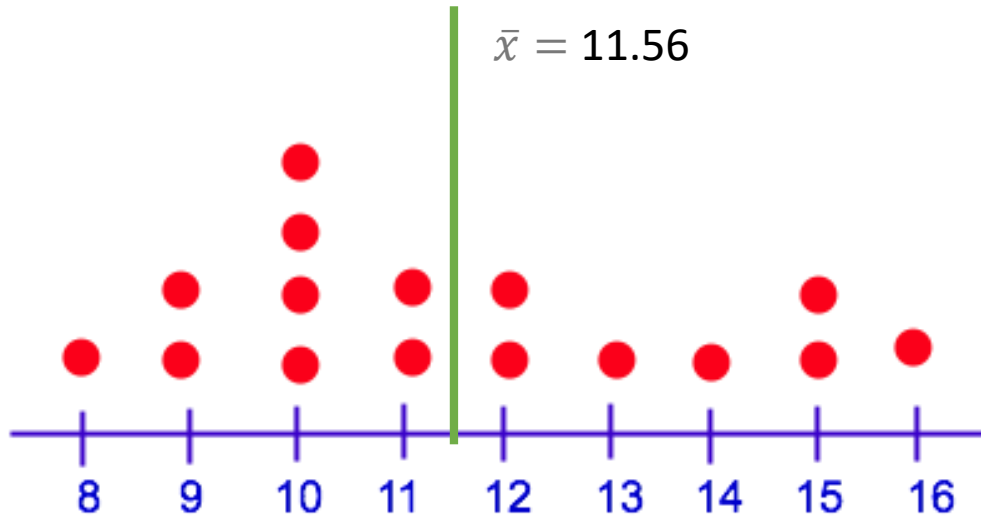
$$RIQ = P_{75} - P_{25} = 2.175 - 0.465$$



# Medidas de Dispersión

## Desviación Estándar

Es el promedio de las distancias de cada observación a la media muestral. Esta medida indica que tan alejados o cercanos están los datos de la muestra al promedio general.



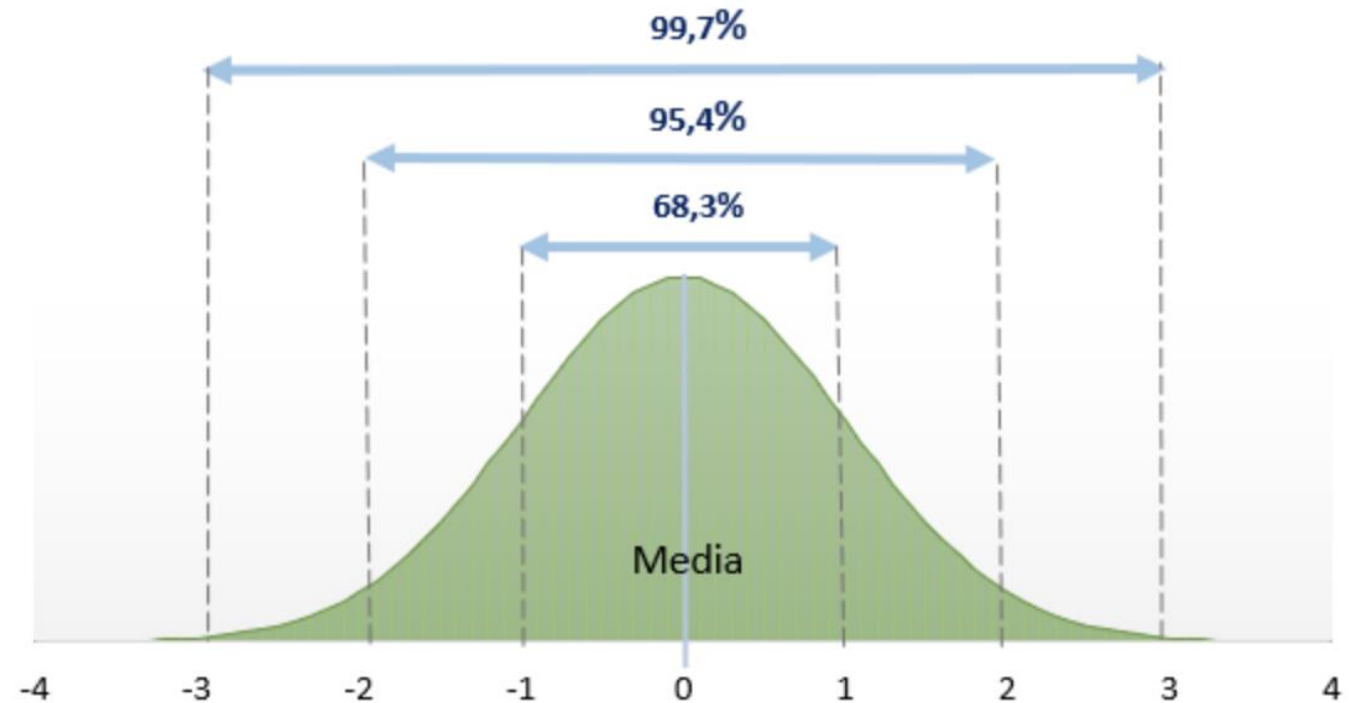
$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$
$$s = \sqrt{\frac{n \sum (x^2) - (\sum x)^2}{n(n - 1)}}$$

# Medidas de Dispersión

## Desviación Estándar

El cálculo de la desviación estándar es de gran importancia ya que según reglas matemáticas se sabe que para muchos conjuntos de datos el 95% de la información se encuentra a dos desviaciones estándar de la media.

Adicionalmente, la inferencia estadística se basa en determinar los valores poblacionales de la desviación estándar y de la varianza (el cuadrado de la desviación, notada como  $\sigma^2$ )



# Medidas de Dispersión

## Desviación Estándar

Aunque se utilizarán herramientas computacionales para el cálculo de las medidas de dispersión, el siguiente es el procedimiento general para calcular la desviación estándar de forma manual

**Paso 1:** Calcular la media.

**Paso 2:** Restar a cada dato el valor de la media

**Paso 3:** Elevar al cuadrado cada valor obtenido en el paso 2

**Paso 4:** sumar los valores que resultaron del paso 3.

**Paso 5:** dividir entre el número de datos menos 1.

**Paso 6:** sacar la raíz cuadrada del paso 4.

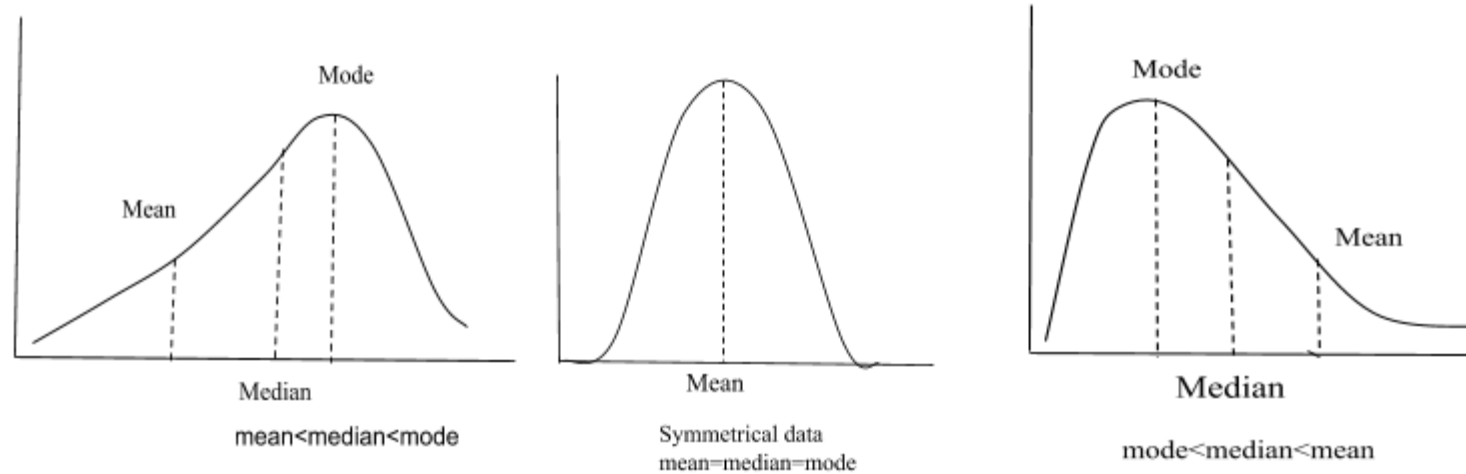
# Medidas de Simetría y Curtosis



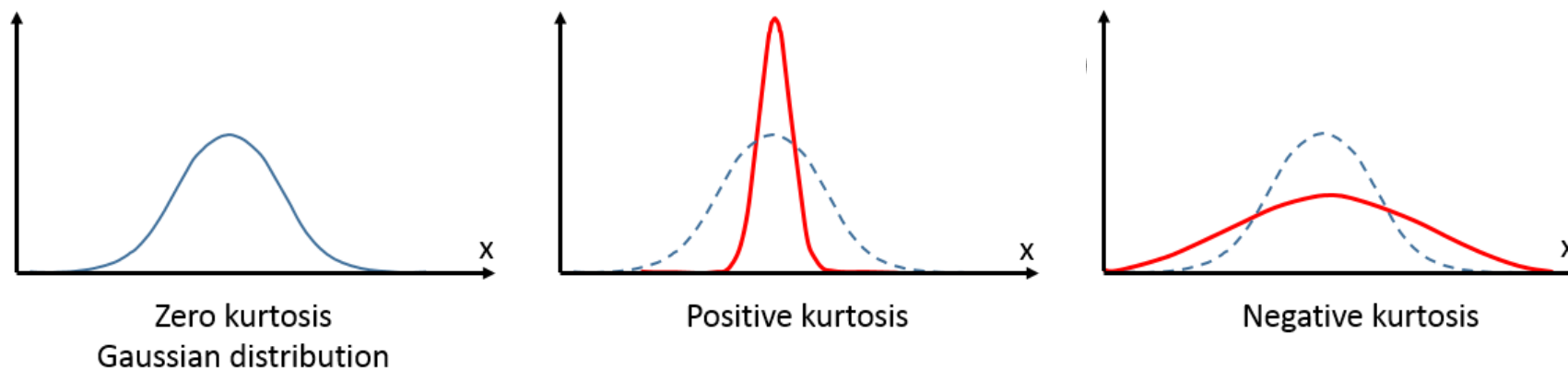
FUNDACIÓN UNIVERSITARIA  
DE CIENCIAS DE LA SALUD  
**FUCS**

# Medidas de Simetría y Curtosis

## Simetría



## Curtosis



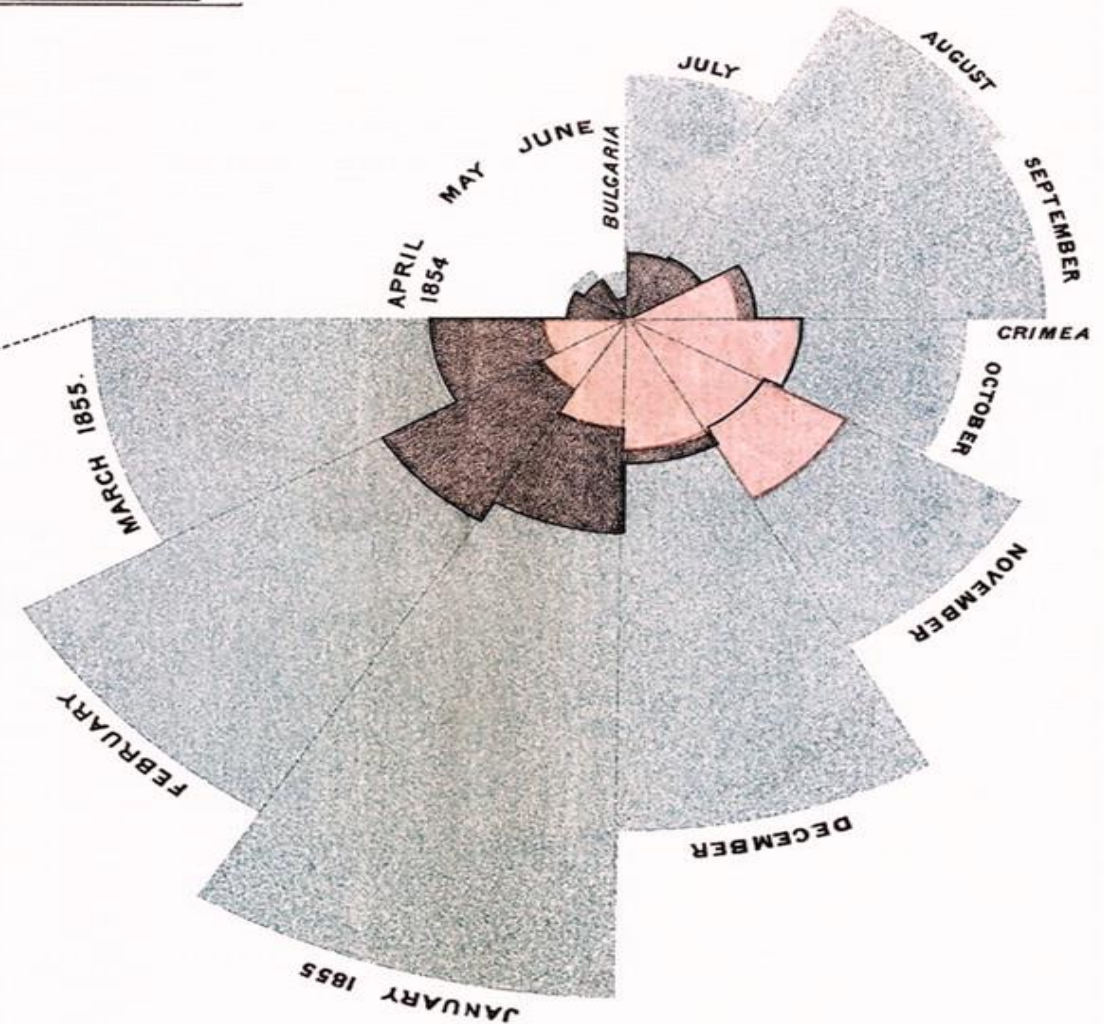
# Visualización de Datos



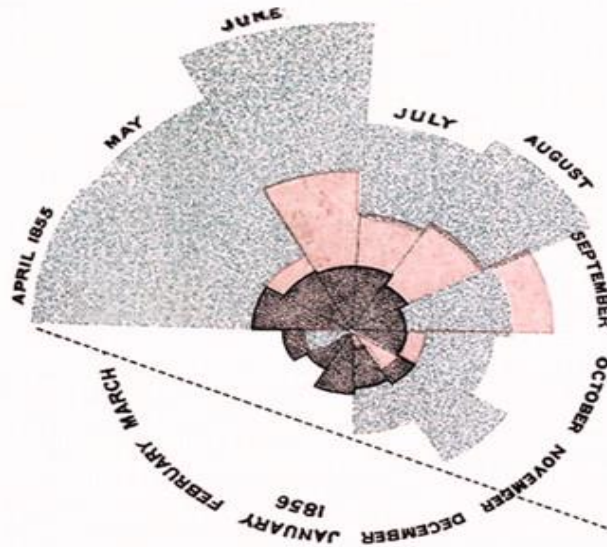
FUNDACIÓN UNIVERSITARIA  
DE CIENCIAS DE LA SALUD  
**FUCS**

## DIAGRAM OF THE CAUSES OF MORTALITY IN THE ARMY IN THE EAST.

1.  
APRIL 1854 to MARCH 1855.



2.  
APRIL 1855 to MARCH 1856.



*The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex.*

*The blue wedges measured from the centre of the circle represent area for area the deaths from Preventable or Mitigable Zymotic diseases; the red wedges measured from the centre the deaths from wounds; & the black wedges measured from the centre the deaths from all other causes.*

*The black line across the red triangle in Nov'r 1854 marks the boundary of the deaths from all other causes during the month.*

*In October 1854, & April 1855; the black area coincides with the red;*

*in January & February 1856, the blue coincides with the black.*

*The entire areas may be compared by following the blue, the red & the black lines enclosing them.*



# Visualización de Datos

Uno de los objetivos de los métodos estadísticos es lograr analizar la información de manera grafica con el fin de que sea mucho mas fácil de interpretar y comunicar. En esta sección se mostrarán algunos ejemplos de visualización de datos las cuales dependen fundamentalmente del tipo de dato a analizar



Tomado de: <https://thevizcorner.wordpress.com/2014/09/06/cataloguing-visualization/>



# Visualización de Datos

## Diagramas de Proporciones

Se usan mayormente para datos categoricos con el fin de representar la participacion de cada categoria en la muestra.

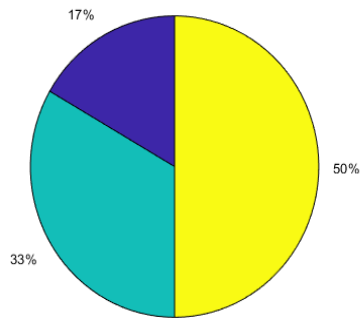


Diagrama de Pastel

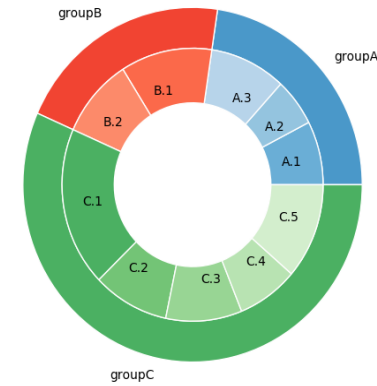


Diagrama de Dona o Sunburst

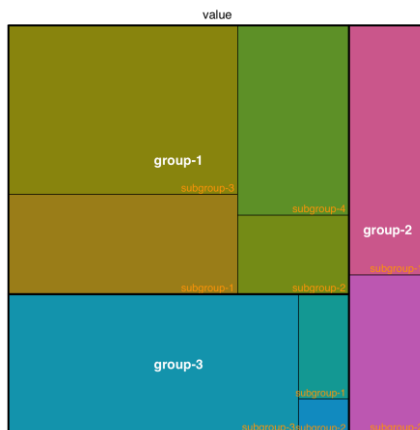
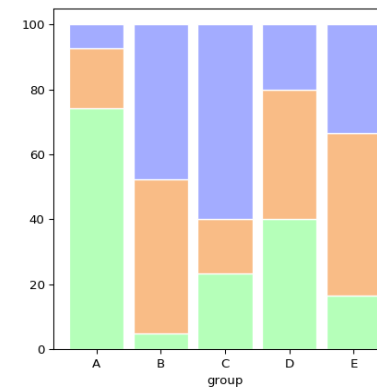


Diagrama Treemap



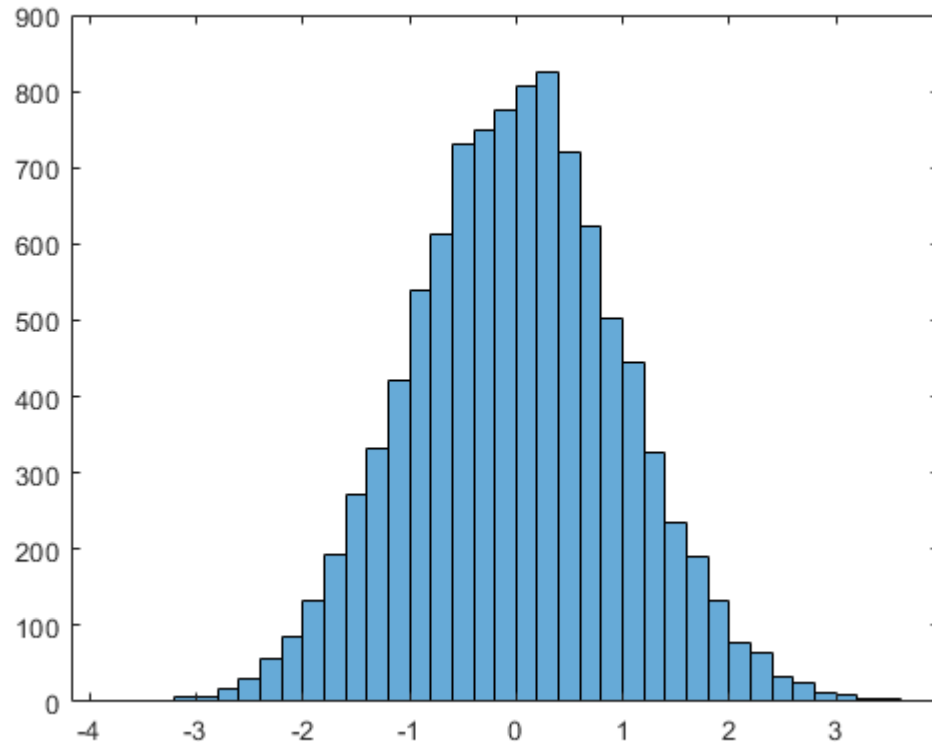
Barras Apiladas

# Visualización de Datos

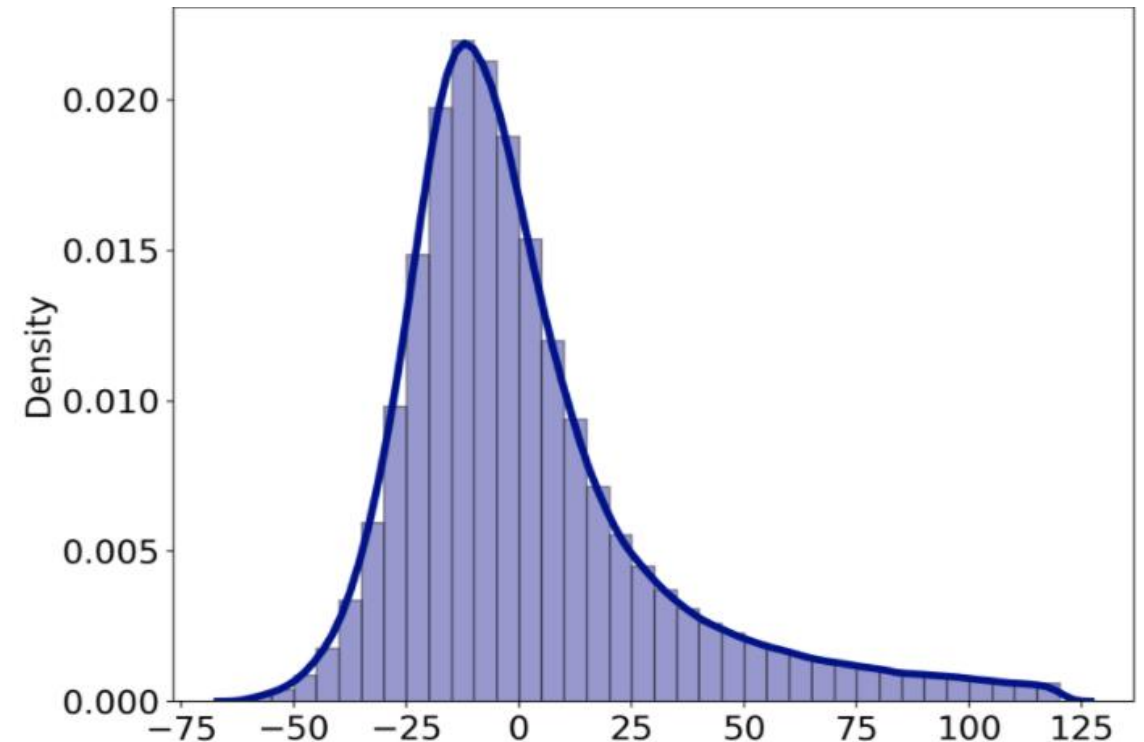
## Histogramas y Densidades

Se usan únicamente para datos numéricos. Resume perfectamente la dispersión, simetría y la curtosis de la distribución, además muestra la existencia de modas y la localización de la media

**Histograma**



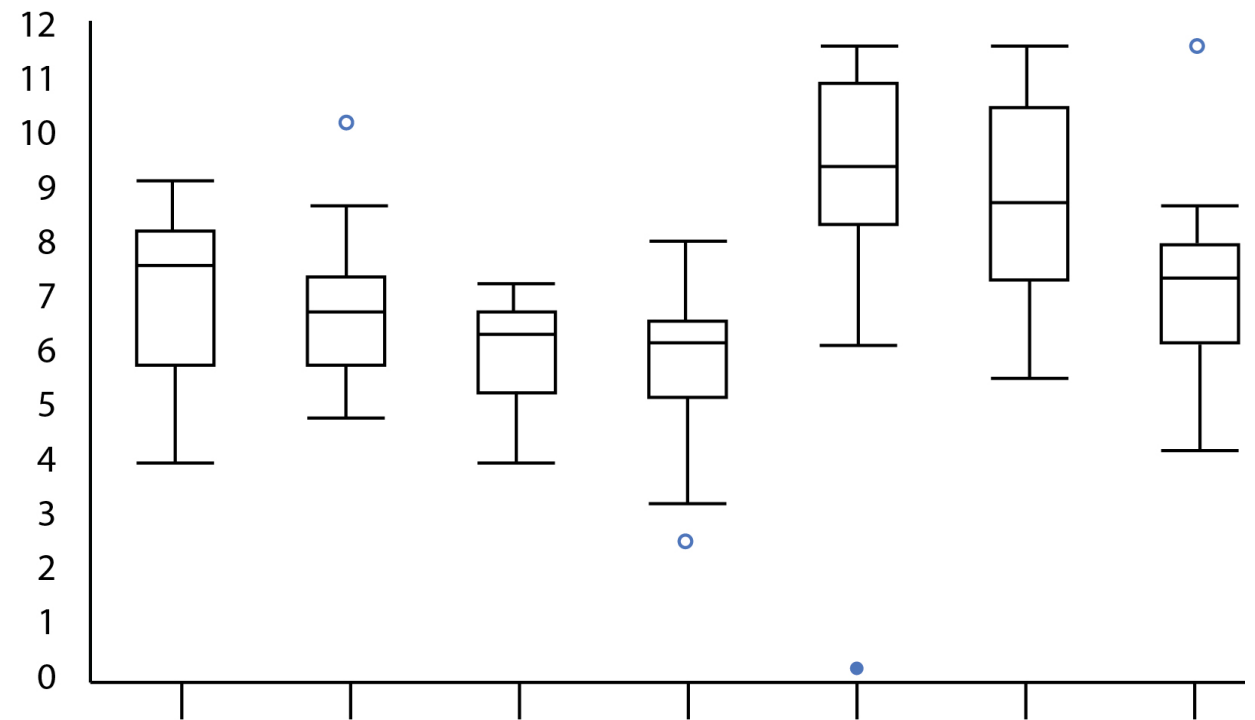
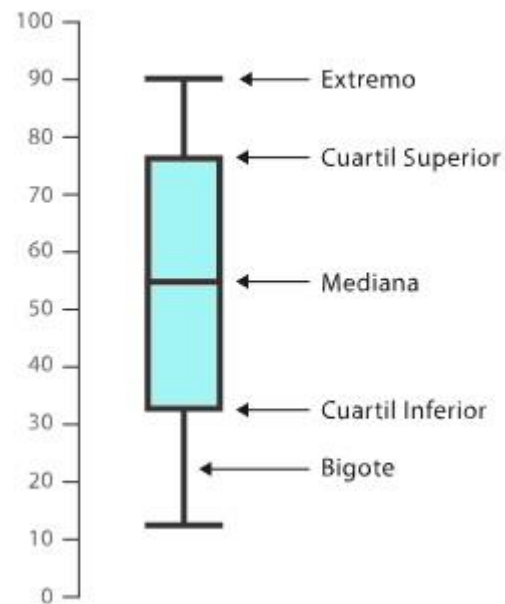
**Densidad**



# Visualización de Datos

## Diagrama de Cajas y Bigotes

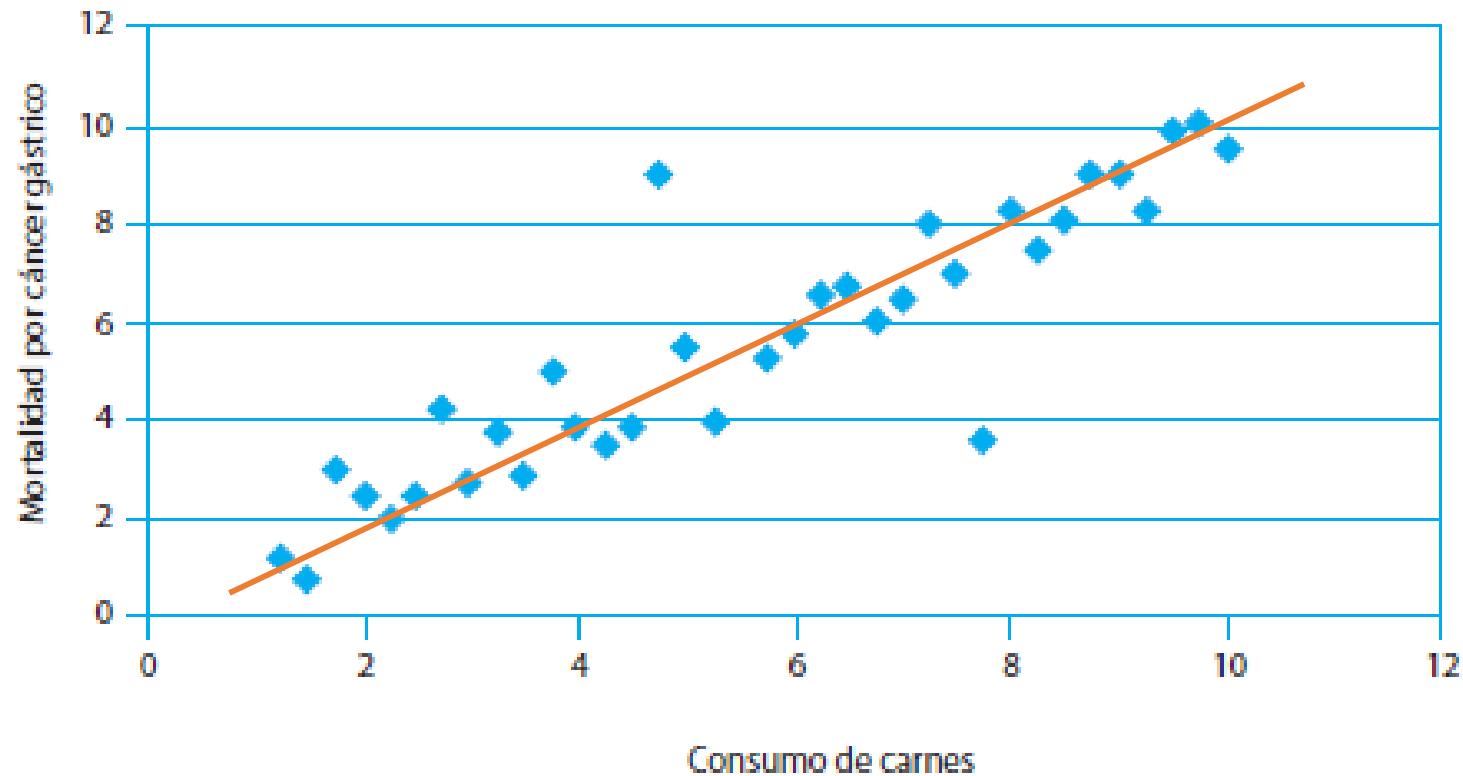
Se usan unicamente para datos numéricos. Resume la dispersión y simetría, además muestra localización de la mediana y la existencia de datos atípicos



# Visualización de Datos

## Diagrama de Dispersión

Se usan únicamente para datos numéricos. Muestra la relación lineal que existen entre dos variables



# Actividad

Se propone, con el fin de afianzar los conceptos descritos durante la sesión identificar la mayor cantidad de elementos en un artículo científico publicado, a saber:

- Tipos de Medida
- Escalas de medición
- Medidas tendencia central
- Medidas de localización
- Gráficos usados y su pertinencia según el tipo de dato

