

Introducción a la Estadística Descriptiva

Ing. Nicolás G Tripp

Algunas definiciones

La Estadística se ocupa de los métodos y procedimientos para recolectar, clasificar, resumir, hallar patrones y analizar datos

- **Individuos:** son las personas u objetos que contienen la información que se quiere analizar.
- **Caracteres:** propiedades de los individuos, pueden ser cualitativas o cuantitativas
- **Modalidades:** Todos los valores posibles de los caracteres.
- **Clases:** conjunto de una o más modalidades.
- **Población:** es el conjunto de individuos con cierta propiedad común.
- **Parámetro:** Característica medible de una población.
- **Muestra:** subconjunto de la población.
- **Estadístico:** Característica medible de una muestra.

Poblaciones

Ejemplos

- Niños nacidos durante el año 2017 (¿en todo el mundo?)
- Niños nacidos en Valle de Uco (¿desde que hay registros?)
- Niños nacidos en Valle de Uco durante el año 2017 (¿hospitales?)
- Niños nacidos en Valle de Uco durante el año 2017 en el hospital X

Carácter	Tipo	Modalidades
Temperatura [C]	Cuantitativa continua	35,0 a 45,0
Semana Epidemiológica	Cuantitativa discreta	1, 2, 3, 4, ...
Grupo sanguíneo	Cualitativa nominal	O, A, B, AB
Prioridad de urgencia	Cualitativa ordinal	Baja, Media, Alta

Ejemplo

Ejemplo

Se proporcionan los pesos (redondeados a libras) de niños nacidos en cierto hospital durante un cierto intervalo de tiempo

4, 8, 4, 6, 8, 6, 7, 7, 7, 8, 10, 9, 7, 6, 10, 8, 5, 9, 6, 3, 7, 6, 4, 7, 6, 9,
7, 4, 7, 6, 8, 8, 9, 11, 8, 7, 10, 8, 5, 7, 7, 6, 5, 10, 8, 9, 7, 5, 6, 5.

Organización de la información

Clasificación Cuando hay muchos individuos o la variable es continua se resume la información mediante “clases”

Las clases se obtienen de la siguiente forma

- 1) **Rango**=máx observado – mínimo observado
- 2) **Número de intervalos:** $k=1+\log(n)/\log(2)$ (Método de Sturges)

K debe ser un número impar

- 3) **Amplitud** de la clase = *rango* / *k*
- 4) Comenzar la primer clase con valor mínimo observado e ir sumando amplitudes hasta alcanzar el valor máximo observado.

Organización de la información

4, 8, 4, 6, 8, 6, 7, 7, 7, 8, 10, 9, 7, 6, 10, 8, 5, 9, 6, 3, 7, 6, 4, 7, 6, 9, 7, 4, 7,
6, 8, 8, 9, 11, 8, 7, 10, 8, 5, 7, 7, 6, 5, 10, 8, 9, 7, 5, 6, 5.

$$\text{Rango} = 11 - 3 = 8$$

$$k = 1 + \log(50) / \log(2) = 6,6. \text{ Elegimos } 7$$

$$\text{Amplitud} = 8 / 7 = 1,143 \text{ lb}$$

C1	C2	C3	C4	C5	C6	C7
(3 a 4,14]	(4,14 a 5,29]	(5,29 a 6,43]	(6,43 a 7,57]	(7,57 a 8,72]	(8,72 a 9,86]	(9,86 a 11)
5	5	9	12	9	5	5

Organización de la información

Tablas estadísticas

Para cada clase se definen:

- Marca de la clase (valor medio del intervalo)
- Frecuencia absoluta (n_i)
- Frecuencia relativa (f_i): $f_i = n_i / n$
- Frecuencia absoluta acumulada (N_i): $N_i = \sum_{\text{desde } 1 \text{ hasta } i} n_j$
- Frecuencia relativa acumulada (F_i): $F_i = N_i / n = \sum_{\text{desde } 1 \text{ hasta } i} f_j$

Se llama distribución de frecuencias al conjunto de clases y frecuencias.

Una tabla estadística presenta de forma ordenada las distribuciones de frecuencias.

Tablas estadísticas

Clase	Marca	n	f	N	F
(3 a 4,14]	3,6	5	0,1	5	0,1
(4,14 a 5,29]	4,7	5	0,1	10	0,2
(5,29 a 6,43]	5,9	9	0,18	19	0,38
(6,43 a 7,57]	7,0	12	0,24	31	0,62
(7,57 a 8,71]	8,1	9	0,18	40	0,8
(8,71 a 9,86]	9,3	5	0,1	45	0,9
(9,86 a 11)	10,4	5	0,1	50	1

Ejemplo

Tabla 2: Descripción de casos confirmados de Fiebre amarilla importados. SE 1 a 26. Argentina.

N° de caso	Edad	Sexo	Residencia
1	27 años	Masculino	Comuna 7 (CABA)
2	22 años	Femenino	General Alvarado (BsAs)
3	45 años	Femenino	Comuna 1 (CABA)
4	69 años	Masculino	Lanus (BsAs)
5	69 años	Masculino	General Roca (Rio Negro)
6	27 años	Femenino	Comuna 11 (CABA)
7	24 años	Masculino	Hurlingham (BsAs)

edad	(22 a 37,7]	(37,7 a 53,4]	(53,4 a 69)]
casos	4	1	2

Clasificación por edad

Sexo	M	F
casos	4	3

Clasificación por sexo

Residencia	CABA	BsAs	Rio Negro
casos	3	3	1

Clasificación por
residencia

Tablas estadísticas

Table
2

Facilities by Bed Size

BED SIZE	NUMBER	PERCENT
≤200	151	20.21
201 - 400	289	38.69
401 - 600	166	22.22
≥600	141	18.88
Total	747	100

Tablas estadísticas

Table
10

Incidents by Age

AGE	NUMBER	PERCENT	DEATHS	CASE FATALITY RATE
<1 year	9,275	1.08	203	2.19
1-4	24,734	2.87	540	2.18
5-9	28,094	3.26	618	2.20
10-14	28,065	3.26	559	1.99
15-19	50,883	5.90	1,541	3.03
20-24	66,103	7.67	2,562	3.88
25-34	107,762	12.50	3,847	3.57
35-44	82,781	9.60	2,776	3.35
45-54	97,233	11.28	3,326	3.42
55-64	101,825	11.81	4,058	3.99
65-74	88,158	10.23	4,265	4.84
75-84	90,960	10.55	6,056	6.66
>84	85,932	9.97	7,418	8.63
NK/NR	83	0.01	56	67.47
Total	861,888	100	37,825	4.39

Tablas estadísticas

Table
11

Incidents and Case Fatality Rate by Age and Gender

AGE CATEGORY	NUMBER (FEMALE)	NUMBER (MALE)	DEATHS (FEMALE)	DEATHS (MALE)	CASE FATALITY RATE (FEMALE)	CASE FATALITY RATE (MALE)
<1 year	4,007	5,264	86	117	2.15	2.22
1-4	10,142	14,585	191	349	1.88	2.39
5-9	11,418	16,668	245	373	2.15	2.24
10-14	8,592	19,469	156	403	1.82	2.07
15-19	14,546	36,327	330	1,209	2.27	3.33
20-24	16,957	49,125	423	2,139	2.49	4.35
25-34	26,790	80,943	677	3,169	2.53	3.92
35-44	21,977	60,793	535	2,241	2.43	3.69
45-54	29,124	68,088	820	2,505	2.82	3.68
55-64	37,658	64,150	1,047	3,011	2.78	4.69
65-74	42,825	45,318	1,430	2,834	3.34	6.25
75-84	54,056	36,889	2,588	3,466	4.79	9.40
>84	58,729	27,189	4,089	3,329	6.96	12.24
NK/NR	10	70	6	49	60.00	70.00
Total	336,831	524,878	12,623	25,194	3.75	4.80

Tablas estadísticas

Table
12

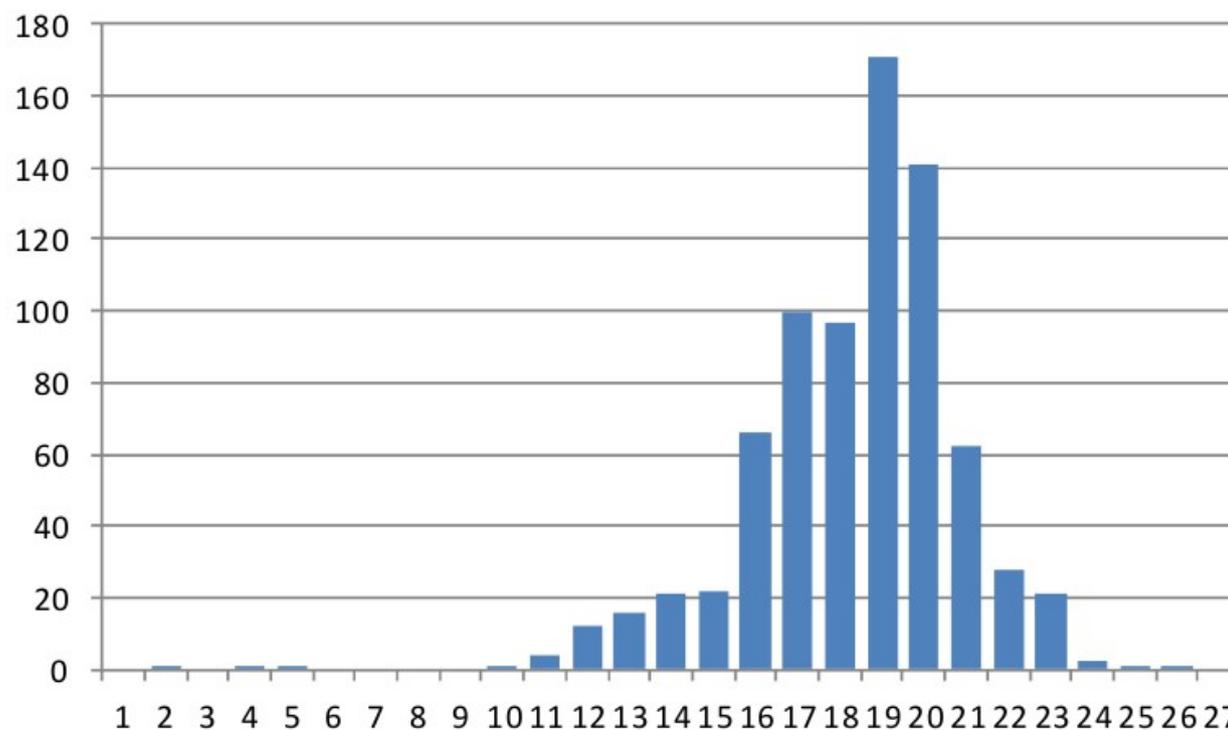
Alcohol Use

ALCOHOL USE	NUMBER	PERCENT
No (confirmed by test)	223,382	25.92
No (not tested)	437,582	50.77
Yes (confirmed by test [beyond legal limit])	84,507	9.80
Yes (confirmed by test [trace levels])	29,789	3.46
Not Applicable	60,390	7.01
NK/NR	26,238	3.04
Total	861,888	100

Presentaciones gráficas de la información

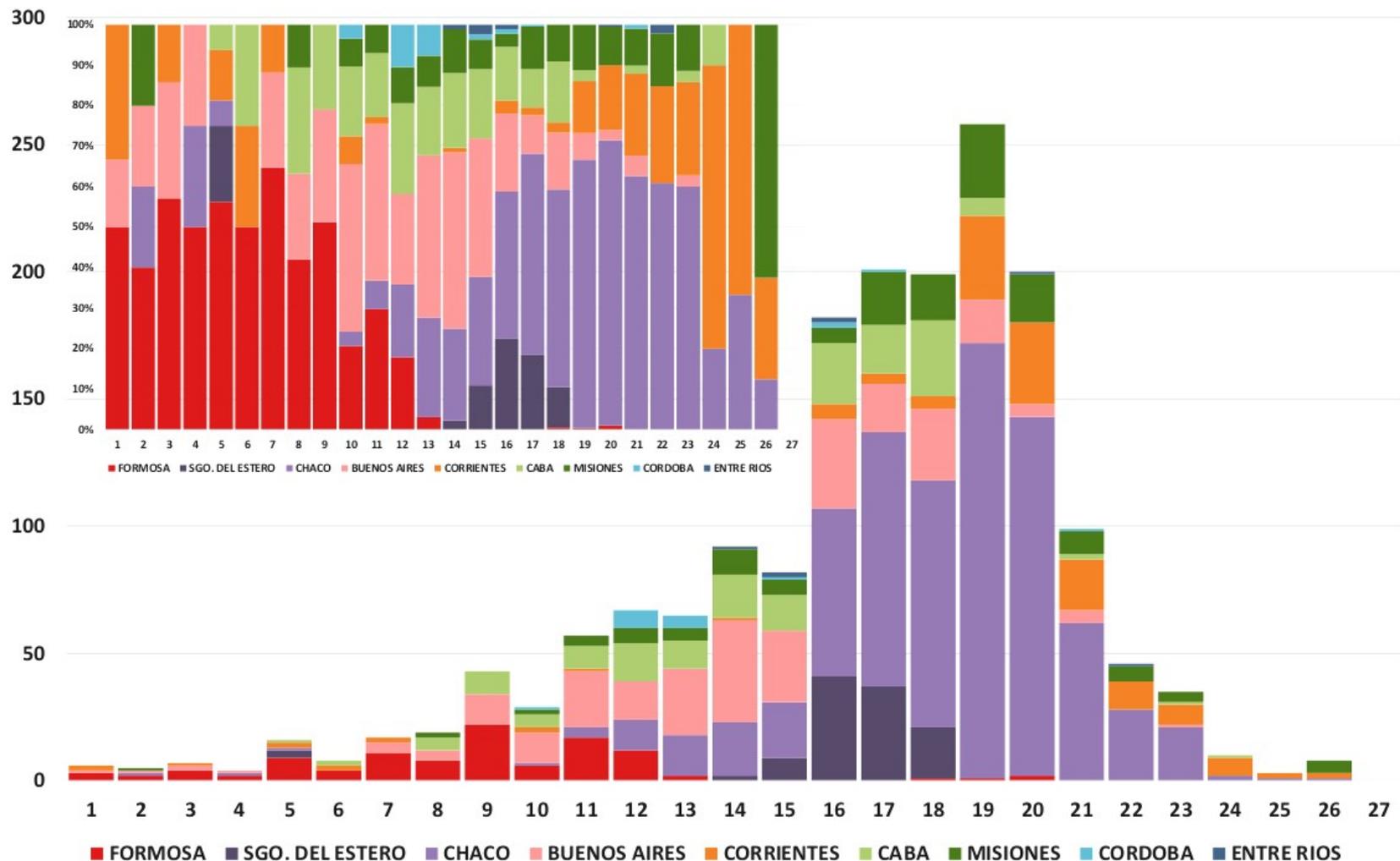
Boletín Integrado de Vigilancia | N° 414- SE 28 2018 | Página 10 de 65

Gráfico 3: CHACO: Casos confirmados y probables de dengue según semana de inicio de síntomas o consulta. SE 1 a 27 de 2018. N=769



Presentaciones gráficas de la información

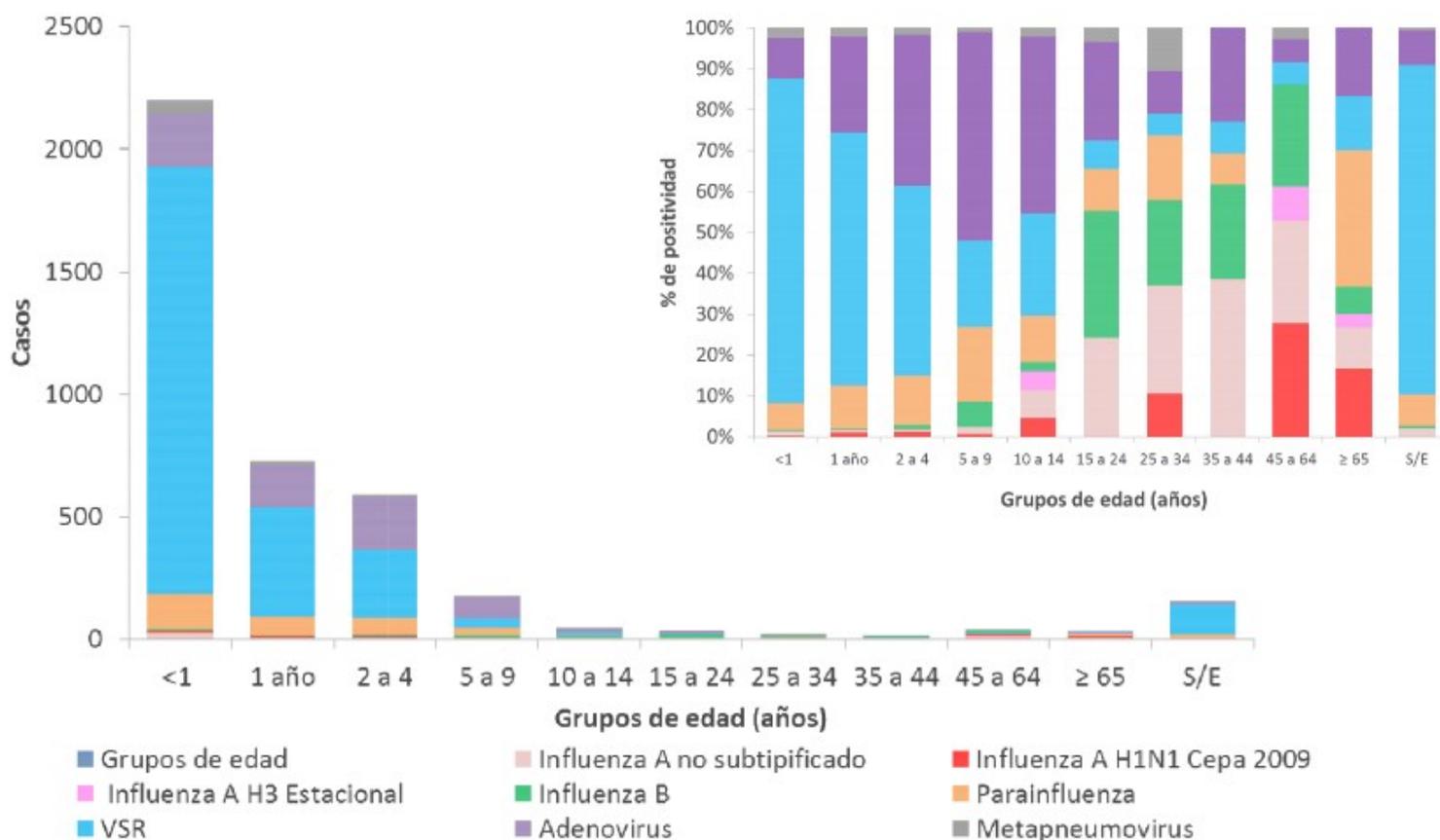
Gráfico 8: ARGENTINA: Frecuencia absoluta y relativa de casos confirmados y probables de dengue por provincia de residencia según semana de inicio de síntomas o consulta. SE 1 a 27 de 2018. N=1758



Fuente: Elaboración propia del Área de Vigilancia de la Salud de la Dirección de Epidemiología en base a información proveniente del Sistema Nacional de Vigilancia de la Salud (SNVS) -Módulos C2 y SIVILA- y del SNVS^{2.0}.

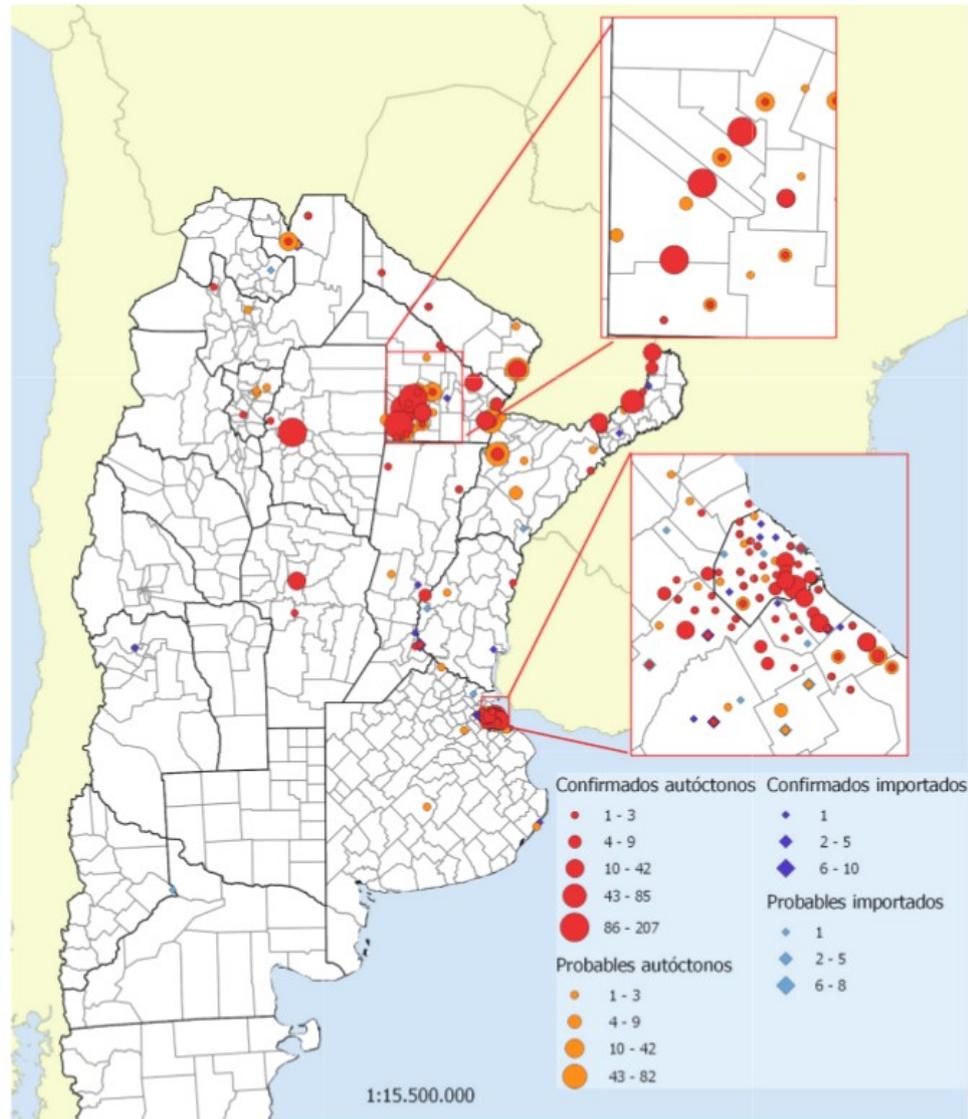
Presentaciones gráficas de la información

Figura 8. Casos de virus respiratorios por grupo de edad. Acumulados a la SE24 de 2018. Argentina.



Presentaciones gráficas de la información

Mapa 1: Casos confirmados y probables de dengue por lugar de residencia según sitio probable de adquisición de la infección. SE 1 a 26 de 2018

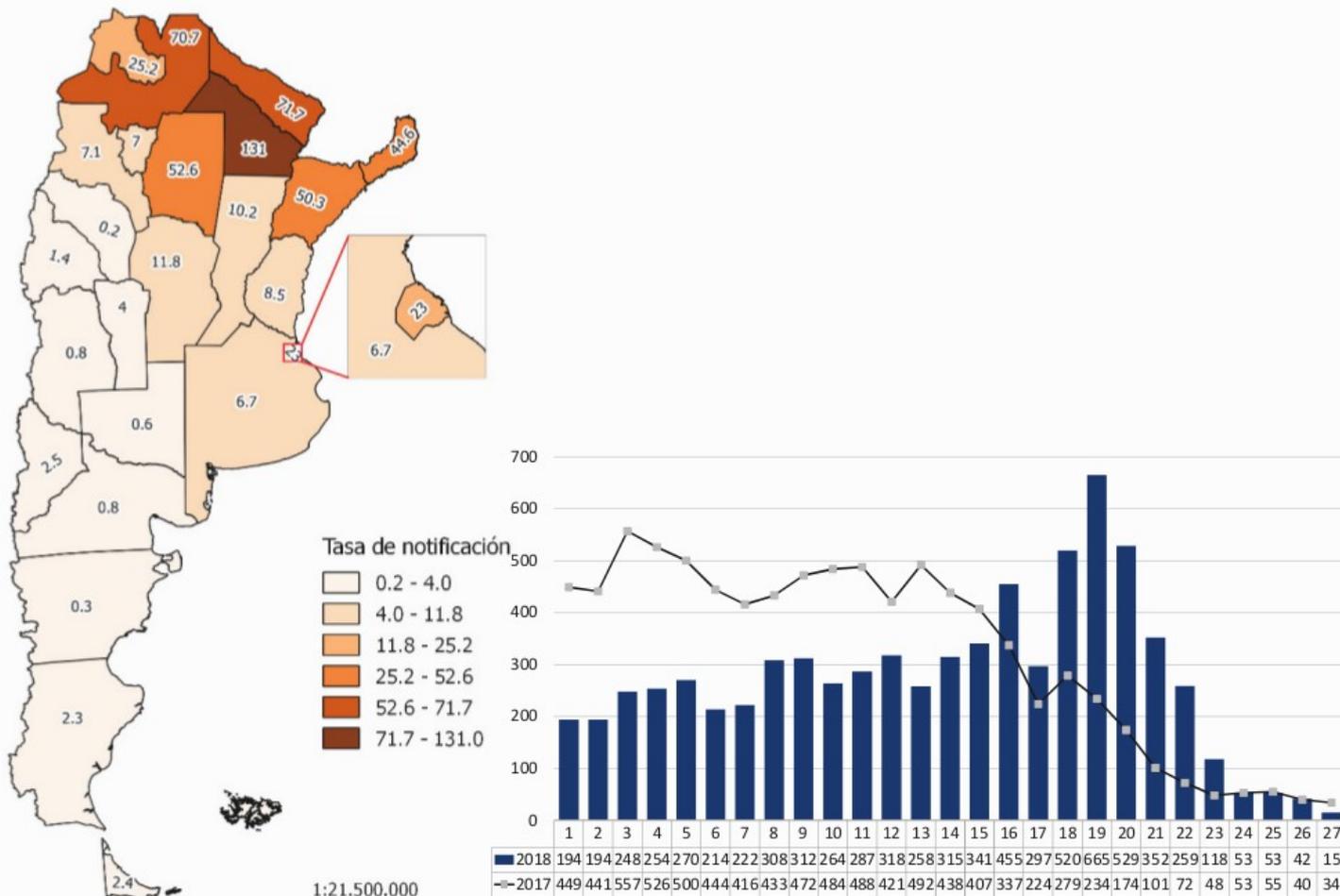


Fuente: Elaboración propia del Área de Vigilancia de la Salud de la Dirección de Epidemiología en base a información proveniente del Sistema Nacional de Vigilancia de la Salud (SNVS) -Módulos C2 y SIVILA- y del SNVS^{2.0}.

Presentaciones gráficas de la información

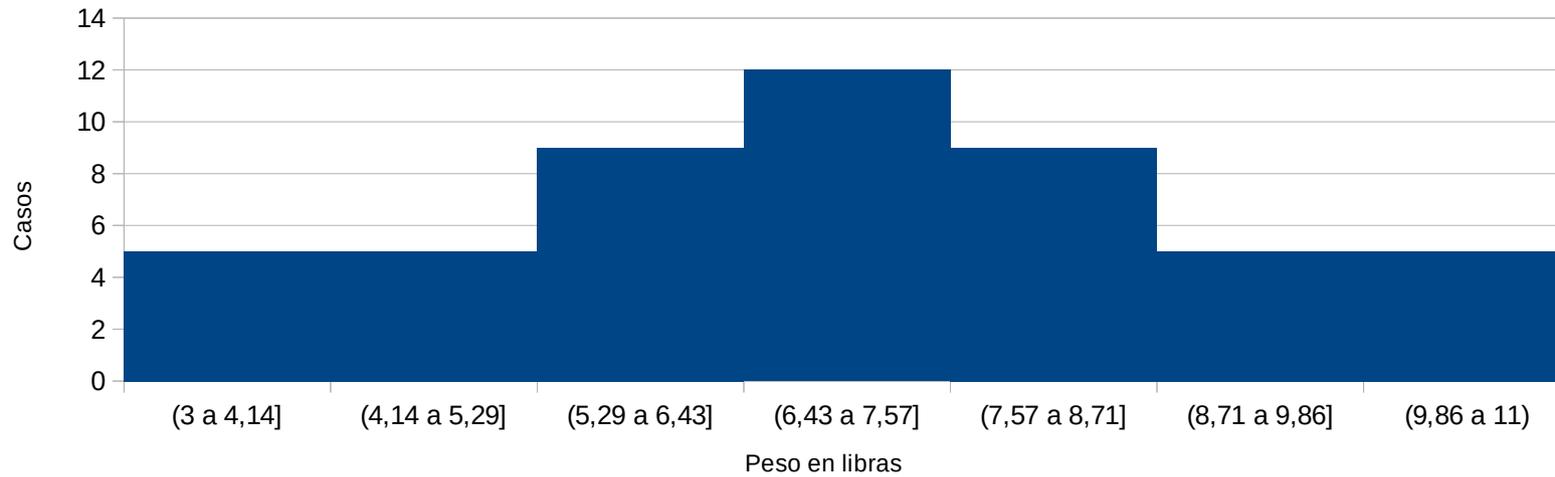
Gráfico 3: Tasa de casos notificados con sospecha de arbovirosis por 100 mil habitantes según provincia de residencia y Casos notificados por semana epidemiológica de inicio de síntomas o consulta. SE 1 a SE 27. Años 2017 y 2018.

Fuente: Elaboración propia del Área de Vigilancia de la Salud de la Dirección de Epidemiología en base a información proveniente

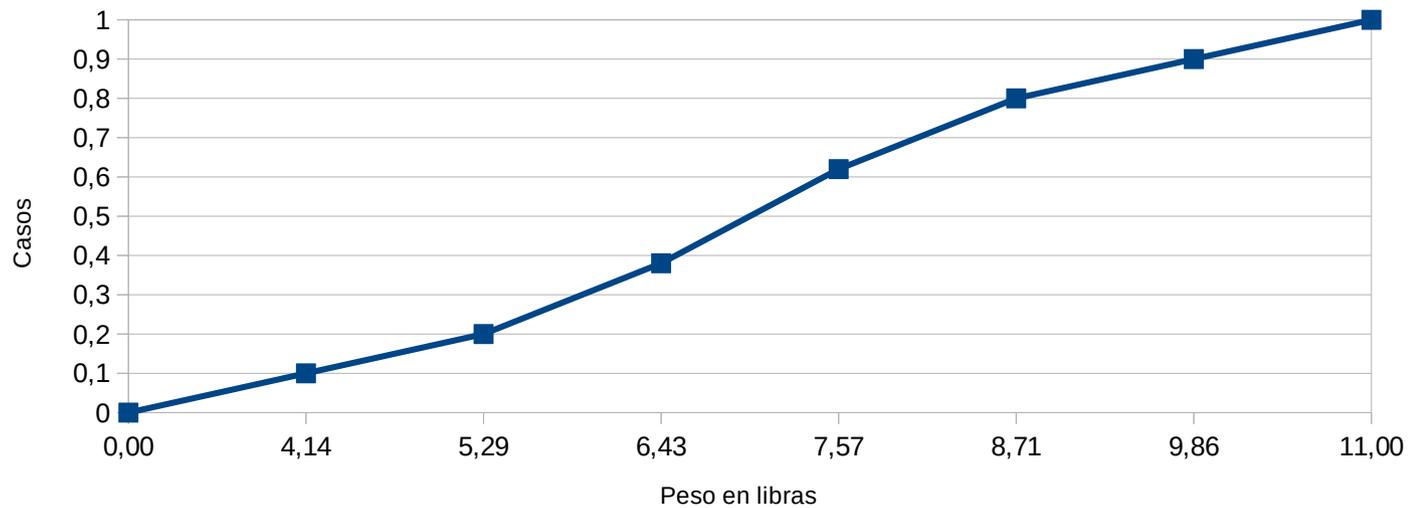


Polígonos e histogramas

Peso de recién nacidos en cierto intervalo de tiempo y lugar



Peso de recién nacidos en cierto intervalo de tiempo y lugar



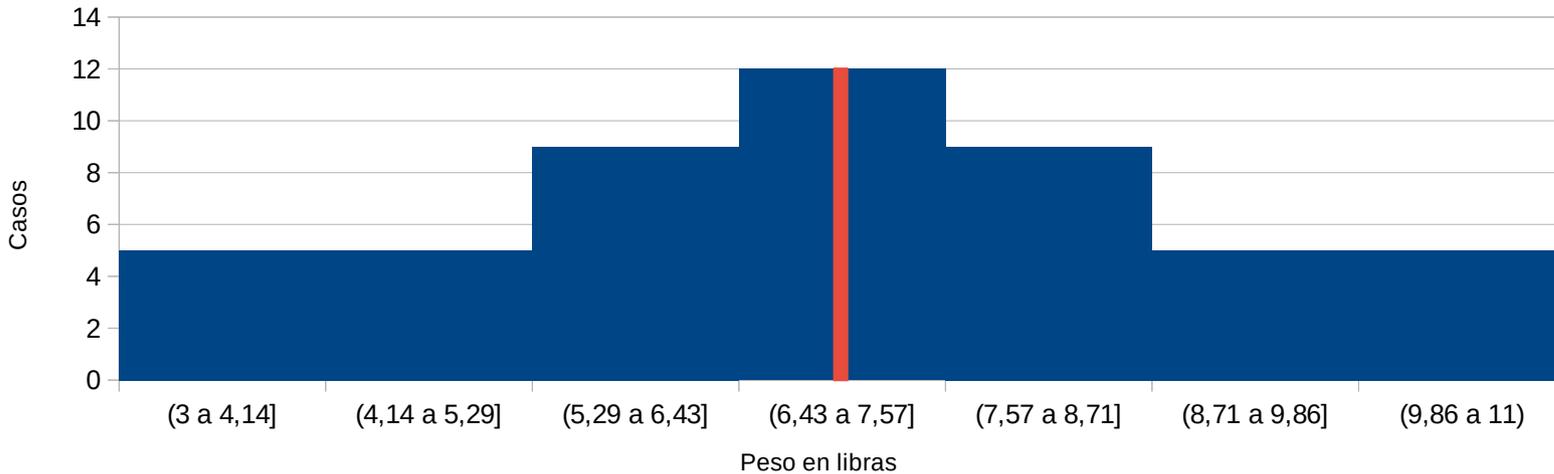
Medidas

Las medidas del conjunto de datos dan una noción rápida de la gráfica de la distribución de frecuencias.

- Medidas de **tendencia central** (dónde se ubica el centro)
- Medidas de **dispersión** (cuán alejados se ubican los extremos)
- Medidas de **asimetría**
- Medidas de **apuntamiento**

Medidas de tendencia central

Peso de recién nacidos en cierto intervalo de tiempo y lugar



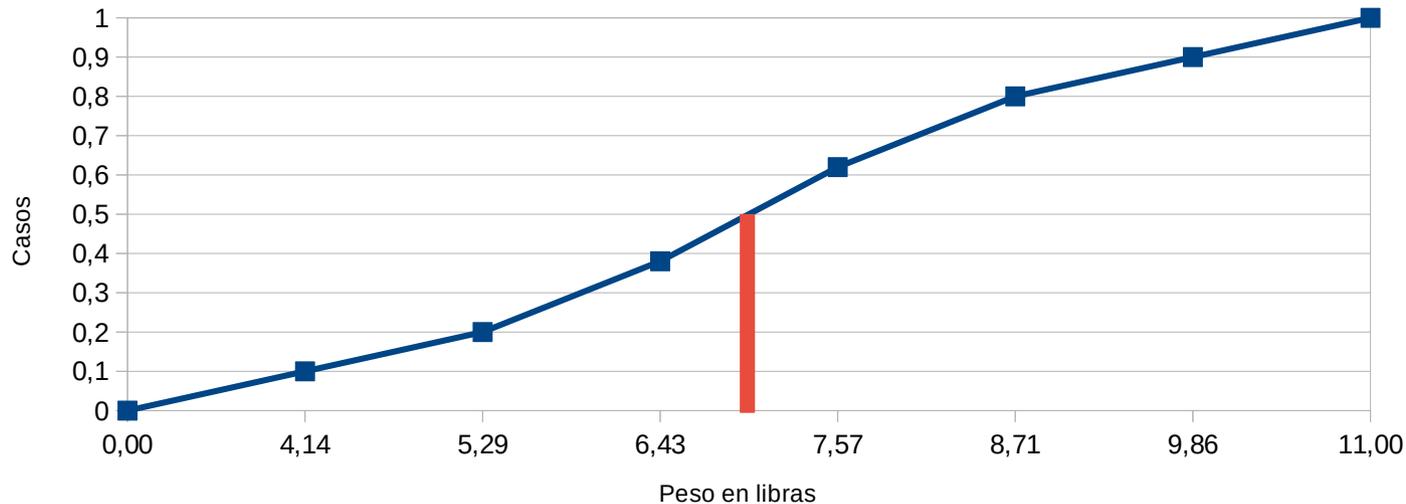
Moda

Dato o intervalo de mayor frecuencia

Media

Promedio aritmético

Peso de recién nacidos en cierto intervalo de tiempo y lugar



Mediana

Divide los datos por la mitad según la frecuencia acumulada

Percentiles

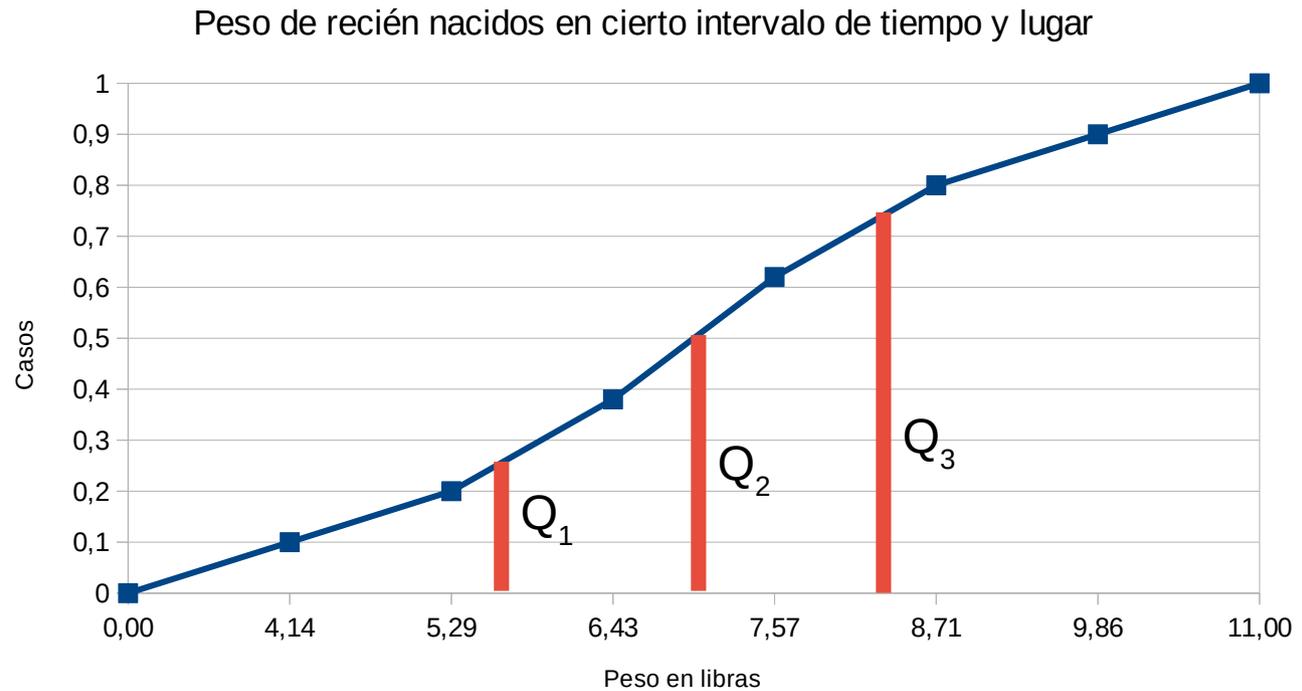
Un intervalo percentil P_α es aquel que contiene un frecuencia acumulada relativa superior a $\alpha\%$

Primer cuartil (Q_1) = P_{25}

Segundo cuartil (Q_2) = P_{50}

Tercer cuartil (Q_3) = P_{75}

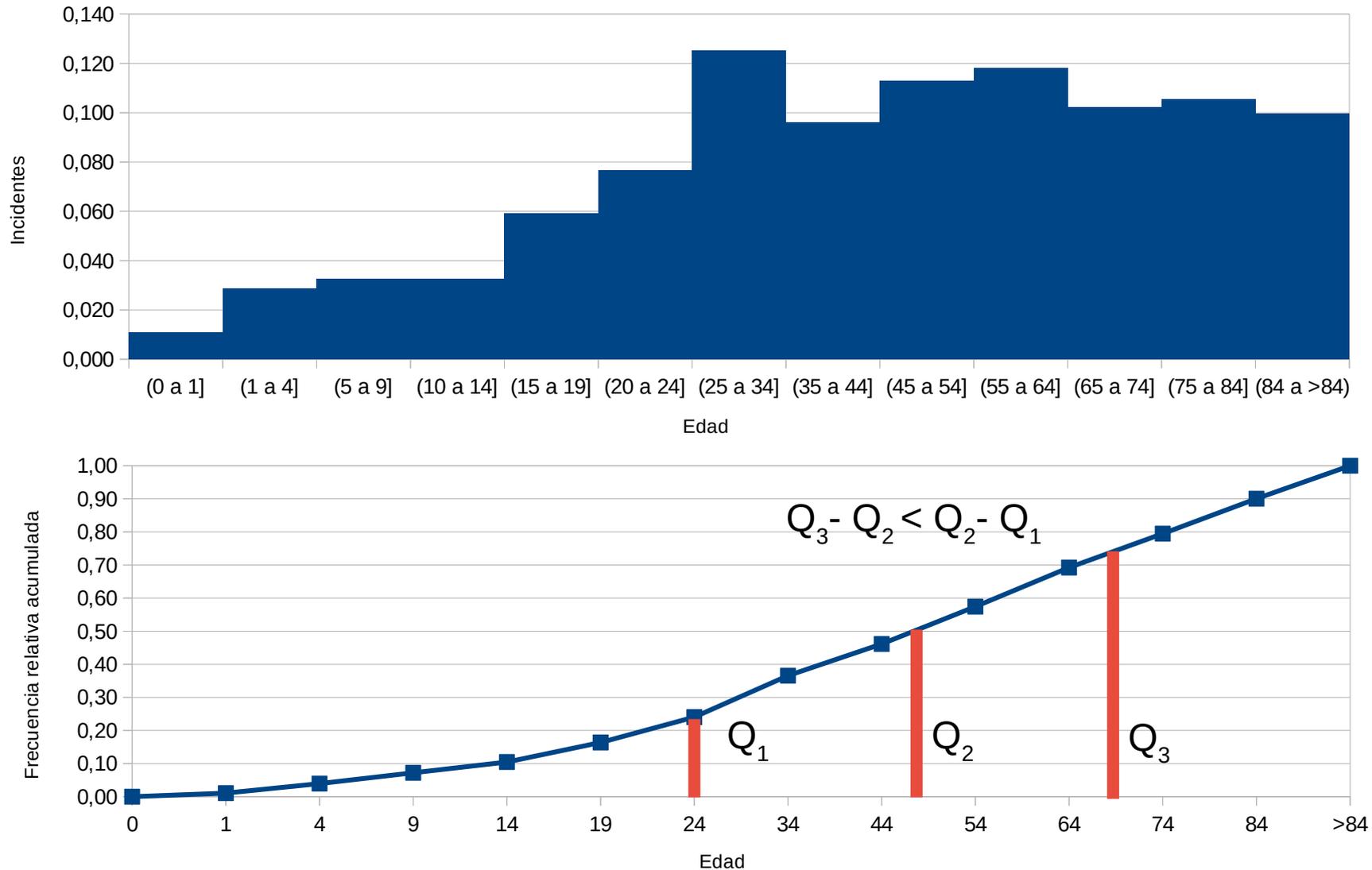
Rango intercuartil = $Q_3 - Q_1$



<http://shiny.chemgrid.org/boxplotr/>

Medidas de asimetría

Incidentes por edad durante 2016, publicacion de colegio de cirujanos norteamericanos



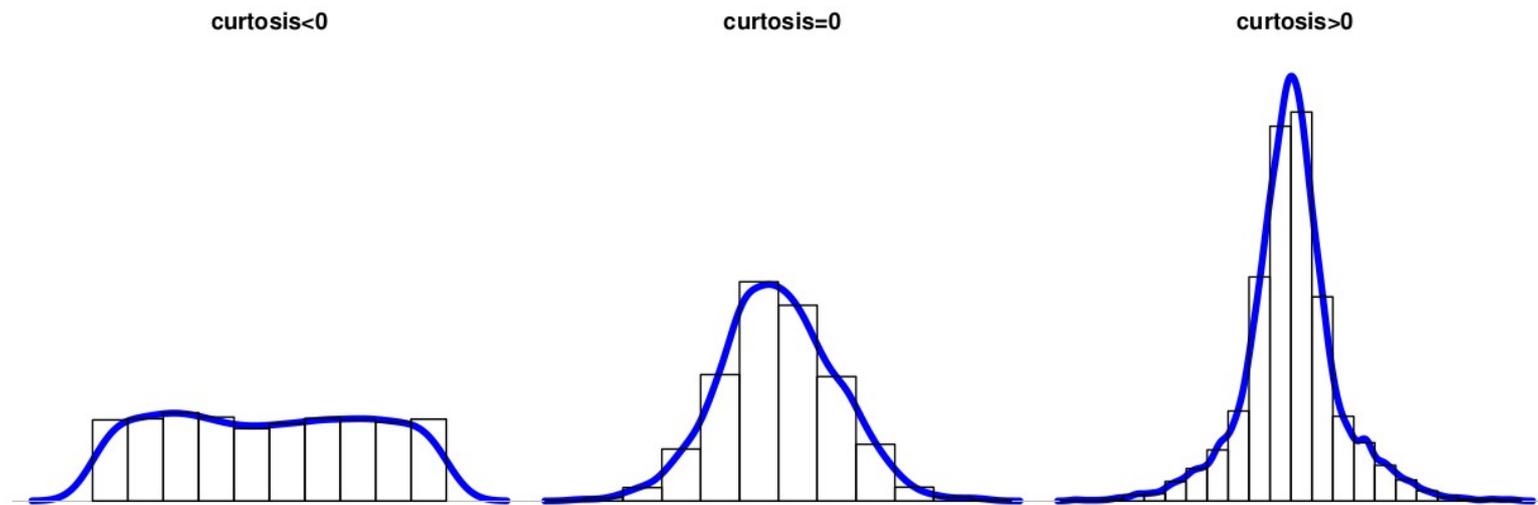
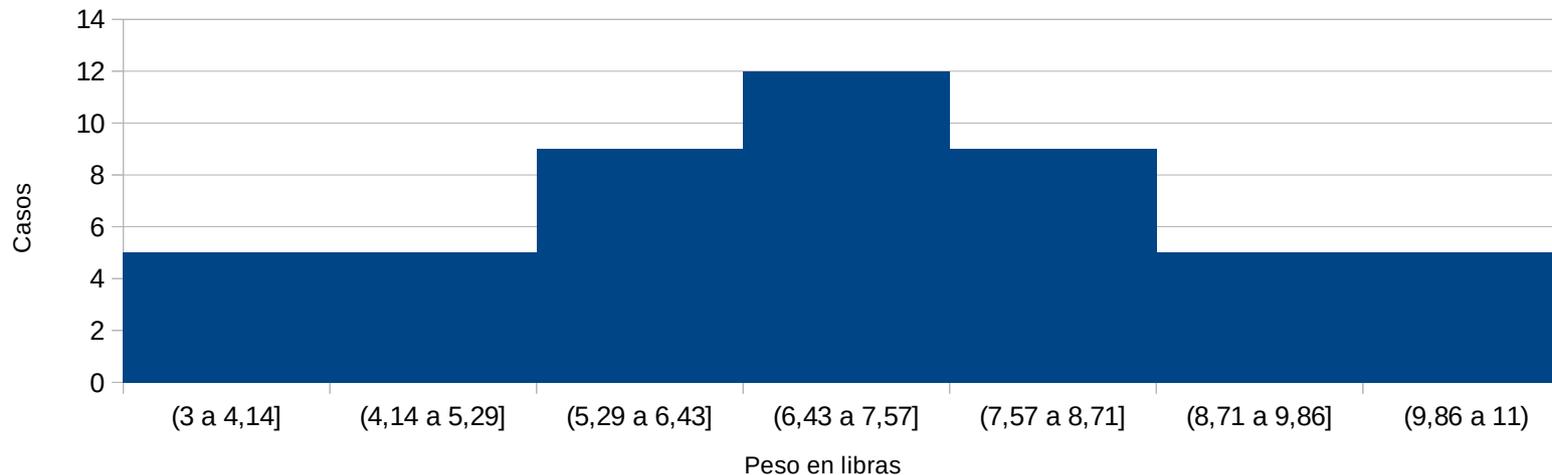


Figura 2.10: Apuntamiento de distribuciones de frecuencias

Medidas de dispersión

Peso de recién nacidos en cierto intervalo de tiempo y lugar



Rango

Varianza

mide la distancia de la observación o intervalo respecto a la media, da unidades cuadráticas.

Desvío estándar

recupera las unidades de medida

Rango	8 lb
Varianza	3,31 lb ²
Desviación estándar	1,82 lb