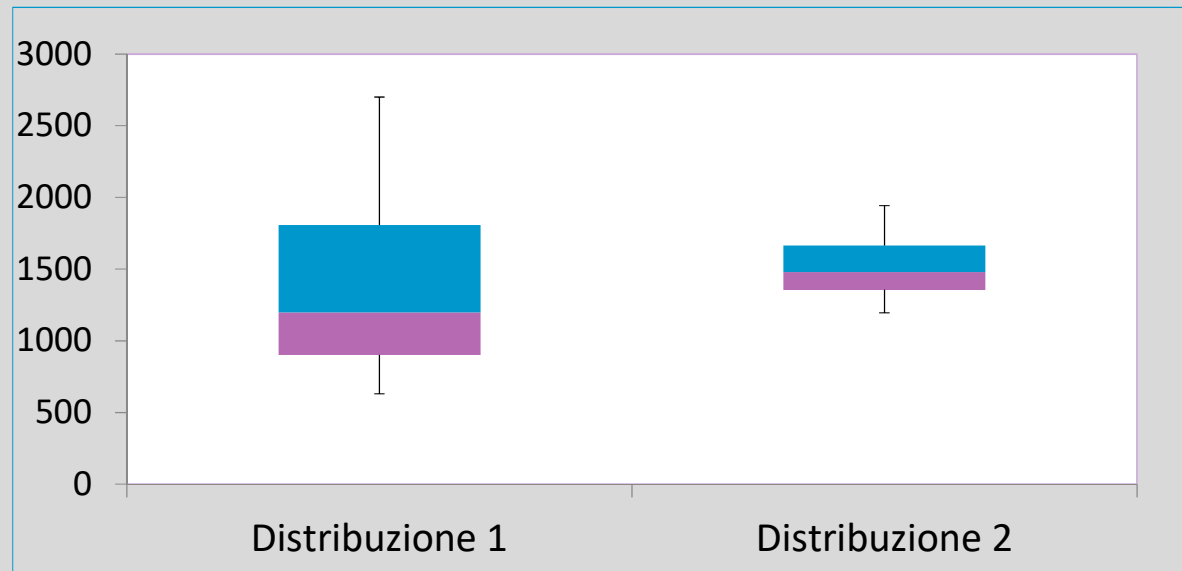
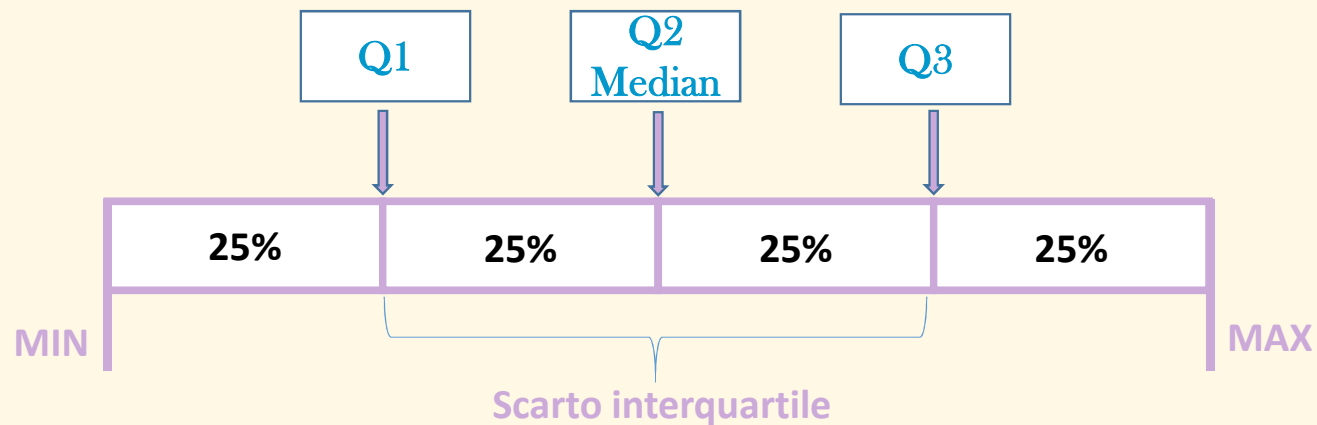


Confronto tra distribuzioni



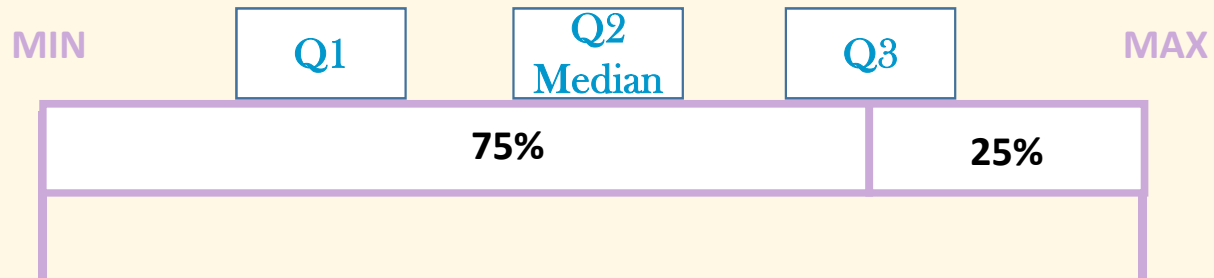
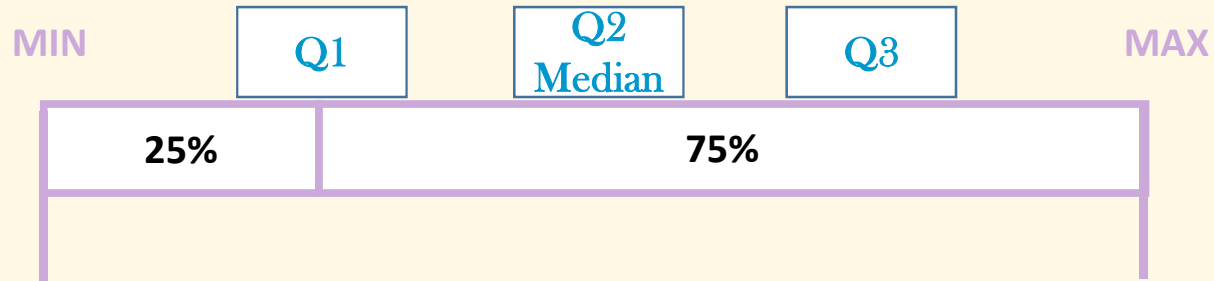
Quartili

I **quartili** sono quei valori che dividono l'insieme dei dati in 4.
La mediana lascia a destra e a sinistra del proprio valore il 50% dei casi, suddividendo la distribuzione in due parti uguali.
Posso dividere in maniera analoga la distribuzione in 4 parti uguali.



Lo **scarto interquartile** si definisce come la differenza tra il terzo e il primo quartile di un insieme di dati.

Quartili

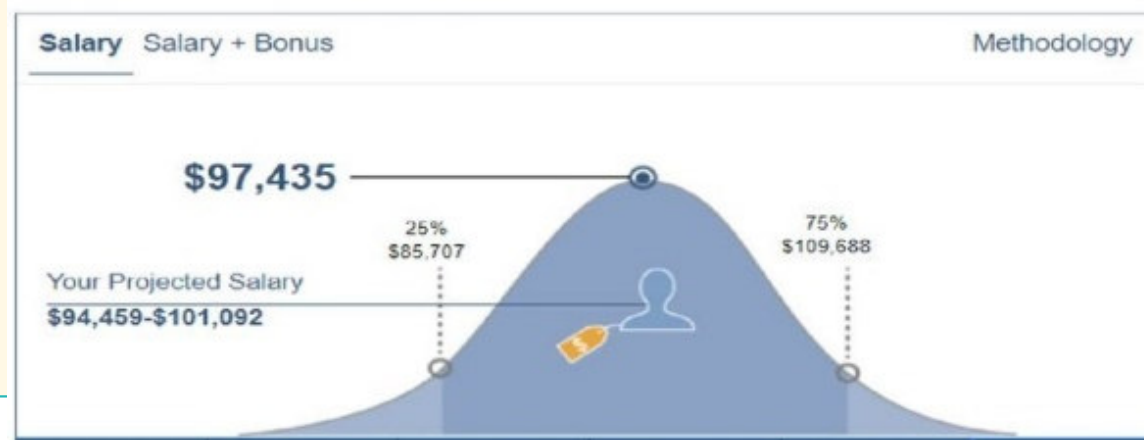


Q3-Q1 = scarto interquartile. É l'ampiezza dell'intervallo che contiene il 50% dei dati posizionati in mezzo alla distribuzione. É una misura di variabilità che supera il problema della dipendenza dai valori estremi.

Utilizzo dei quartili

Esempio: Per un manager è utile sapere che la propria azienda si trova nel primo quartile, se questo corrisponde a un fatturato di 20mln. Vuol dire che fa parte del gruppo del 25% di aziende che ha avuto un fatturato fino a 20mln, mentre il restante 75% ha avuto un fatturato maggiore di 20mln.

Esempio: Si può valutare il posizionamento del proprio salario rispetto alla distribuzione dei salari di altri dipendenti di altre aziende.



Esempio

Calcolo della mediana, dei quartili e dello scarto interquartile per i seguenti salari (N 12)

3310	3355	3450	3480	3480	3490	3520	3540	3550	3650	3730	3925
------	------	------	------	------	------	------	------	------	------	------	------

$(3450+3480)/2 = 3465$
Q1

$(3490+3520)/2 = 3505$
Q2

$(3550+3650)/2 = 3600$
Q3

$Q2 = 0,5 * 12 = 6 \rightarrow$ tra 6 e 7

$Q1 = 0,25 * 12 = 3 \rightarrow$ tra 3 e 4

$Q3 = 0,75 * 12 = 9 \rightarrow$ tra 9 e 10

$IQR = 3600 - 3465 = 135$


Come individuare un outlier

$$< Q1 - 1,5 (IQR)$$

$$= 3465 - 1,5(135) = > 3262,5$$

$$> Q3 + 1,5 (IQR)$$

$$= 3600 + 1,5(135) = < 3802,5$$



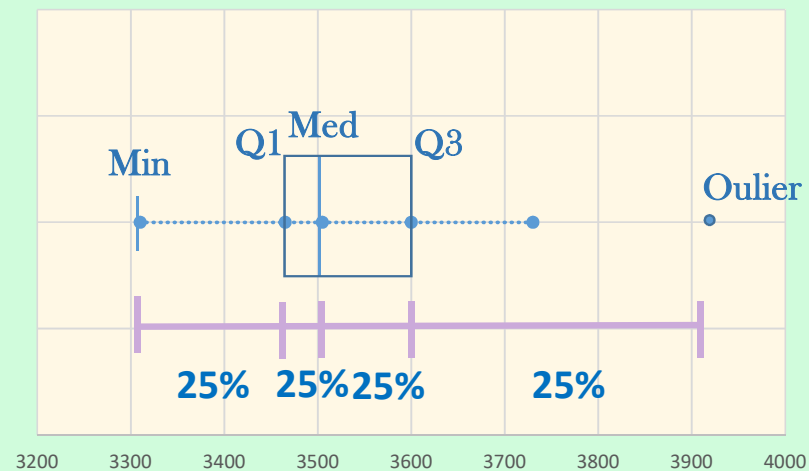
3310	3355	3450	3480	3480	3490	3520	3540	3550	3650	3730	3925
------	------	------	------	------	------	------	------	------	------	------	------

Boxplot

Il diagramma a scatole e baffi, o **boxplot**, è un tipo di rappresentazione grafica adatta a studiare la forma di una distribuzione.

Mostra la mediana, i quartili superiore e inferiore, i valori minimo e massimo ed eventuali outlier nel dataset.

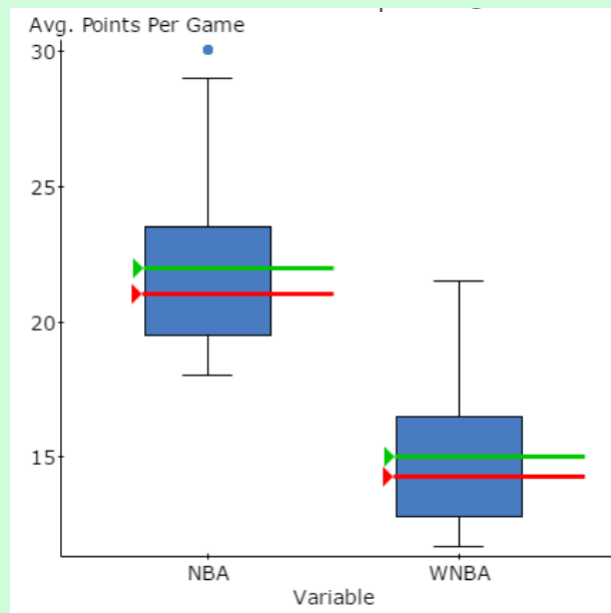
5 numeri EDA		
1	MIN	3310
2	Q1	3465
3	Q2 (mediana)	3505
4	Q3	3600
5	MAX	3925



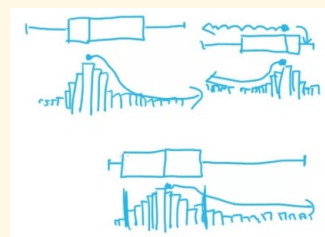
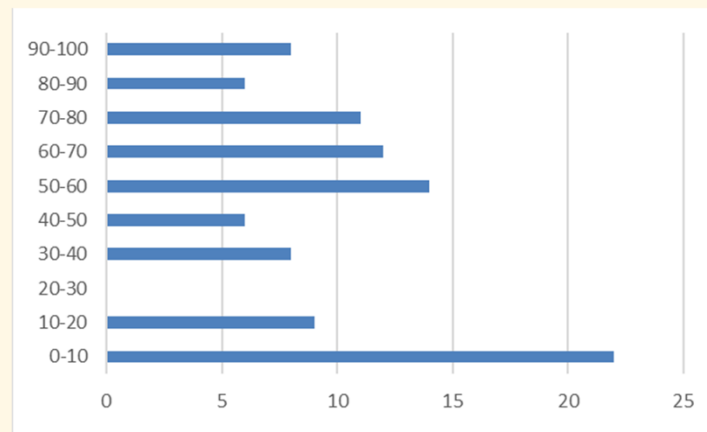
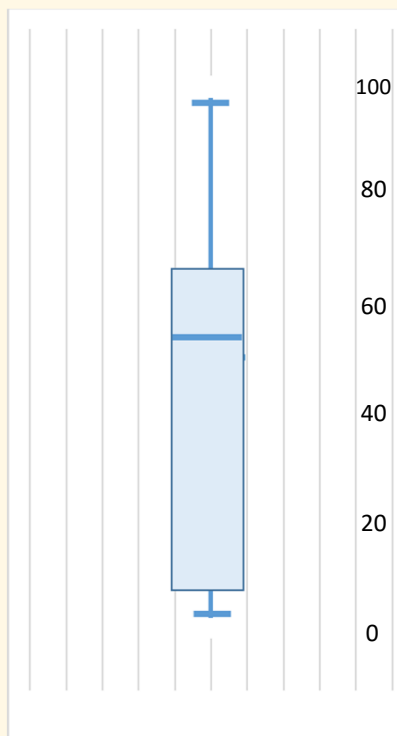
Domande

Se la lunghezza del baffo è corta, cosa ci dice dei valori dei dati?

Se la lunghezza del baffo è lunga, cosa ci dice dei valori dei dati?

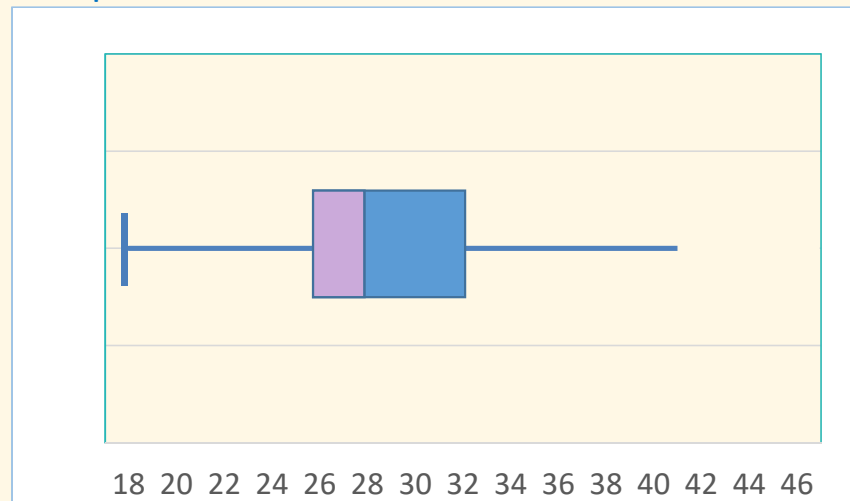


Bloxplot e istogramma a confronto



Domande 1

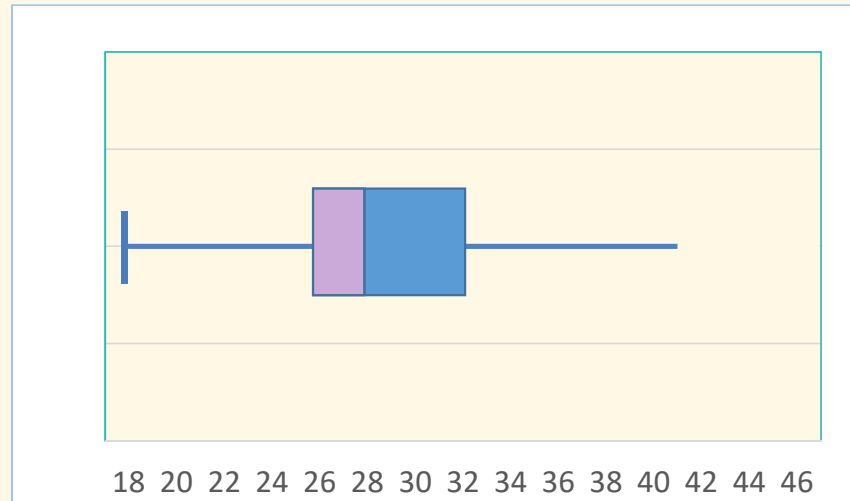
Boxplot – età al matrimonio



- 1- Qual è l'età mediana al matrimonio?
- 2- Vengono celebrati più matrimoni nella fascia d'età 28-32 che 26-28?
- 3- Qual è il valore dell'IQR?
- 4- È possibile individuare un outlier? Come?
- 5- È possibile calcolare il numero di matrimoni che sono stati celebrati tra i 26-28 anni?

Domande 2

Boxplot – età al matrimonio



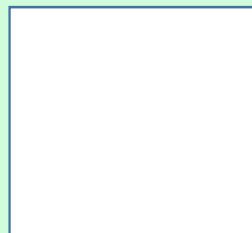
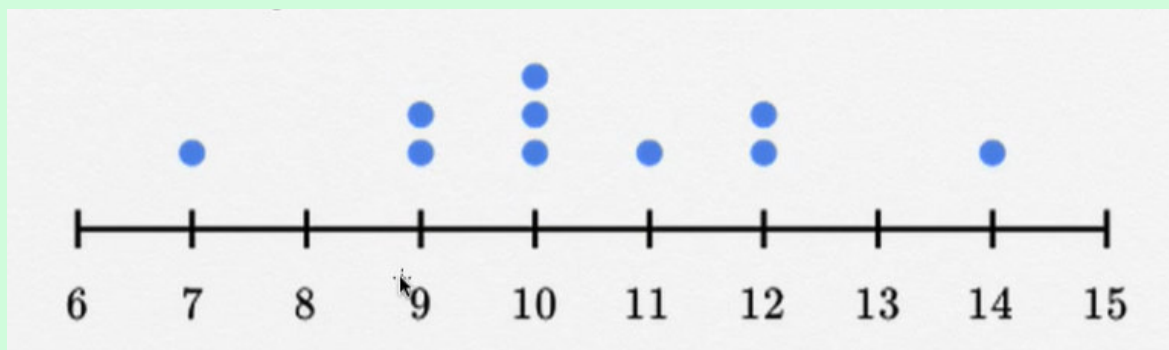
6- Quale percentuale dei matrimoni è celebrata a un'età superiore ai 28 anni?

7- Si può determinare se qualche matrimonio è stato celebrato all'età di 17 anni?

8- In quale fascia d'età si rileva una maggiore variabilità?

Esercizio

Calcolo della mediana, dei quartili e dello scarto interquartile per la distribuzione:



Svolgimento

Scarto interquartile

Modalità	Frequenze assolute	Frequenze cumulate
7	1	1
9	2	3
10	3	6
11	1	7
12	2	9
14	1	10



$$Q2 = 0,5 * 10 = 5 \text{ tra } 5 \text{ e } 6$$

$$Q1 = 0,25 * 10 = 2,5 \text{ quindi } 3$$

$$Q3 = 0,75 * 10 = 7,5 \text{ quindi } 8$$

$$IQR = 12 - 9 = 3$$

Esercizio

Il **boxplot** è un grafico in grado di sintetizzare una distribuzione statistica rispetto a una certa variabile attraverso 5 indici:

x_{\min} Q1 Me Q3 x_{\max}

Estremi della scatola: Q1 e Q3

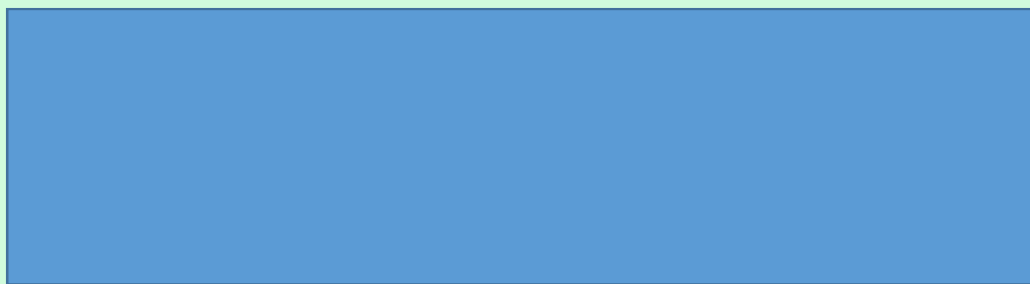
Ampiezza della scatola: Scarto interquartile

Linea all'interno della scatola: Mediana

I segmenti alla fine dei baffi: Estremi della distribuzione x_{\min} e x_{\max}

Determinare i 5 indici per la seguente distribuzione:

3, 5, 5, 5, 6, 8, 8, 12



Esercizio

Il **boxplot** è un grafico in grado di sintetizzare una distribuzione statistica rispetto a una certa variabile attraverso 5 indici:

x_{\min} Q1 Me Q3 x_{\max}

Estremi della scatola: Q1 e Q3

Ampiezza della scatola: Scarto interquartile

Linea all'interno della scatola: Mediana

I segmenti alla fine dei baffi: Estremi della distribuzione x_{\min} e x_{\max}

Determinare i 5 indici per la seguente distribuzione:

3, 5, 5, 5, 6, 8, 8, 12

$x_{\min}= 3$

$x_{\max}=12$

$Q1= 0,25 \cdot 8 = 2$ quindi media quindi tra 2 e 3 = $(5+5)/2$

$Q3= 0,75 \cdot 8 = 6$ quindi tra 6 e 7 = $(8+8)/2$

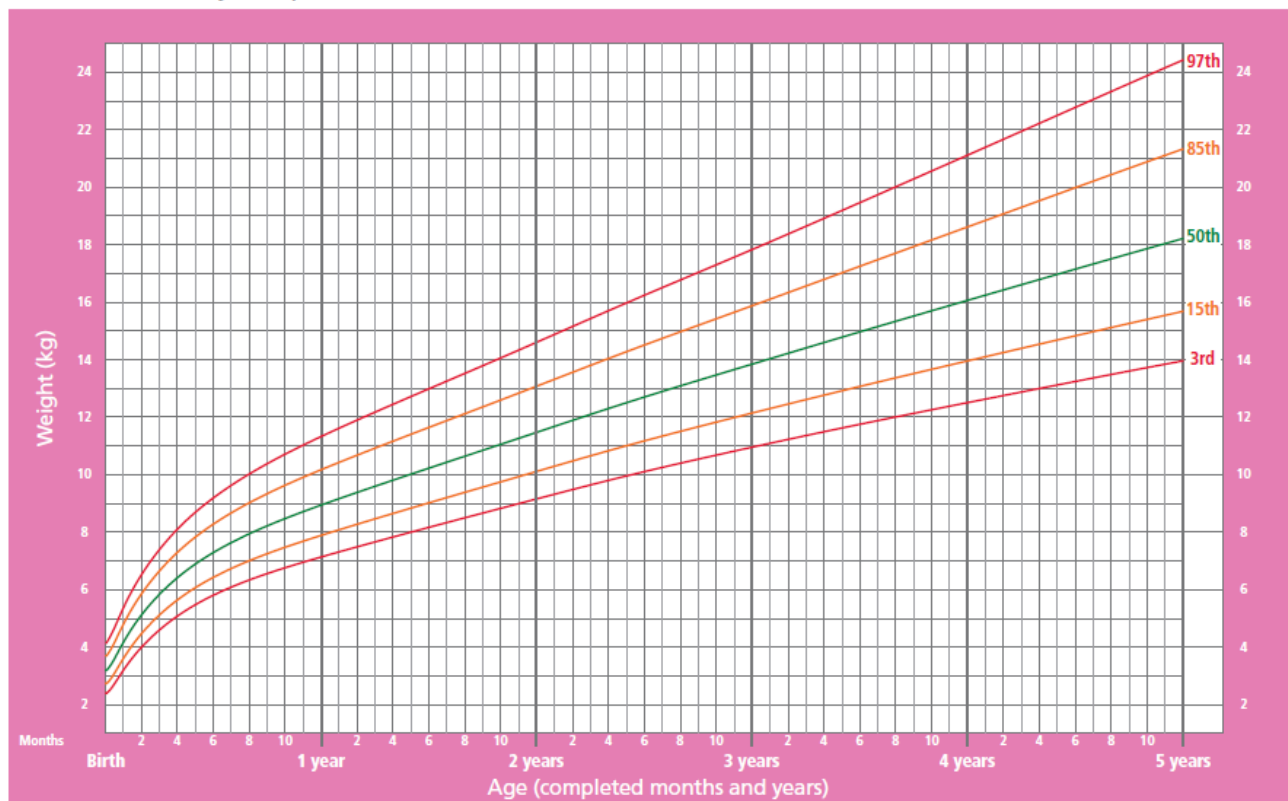
$Me= 0,5 \cdot 8 = 4$ quindi tra 4 e 5 = $(5+6)/2$

Percentili

Altro tipo di quantili sono i percentili. Sono 99 valori della distribuzione che tagliano la stessa in 100 parti uguali.

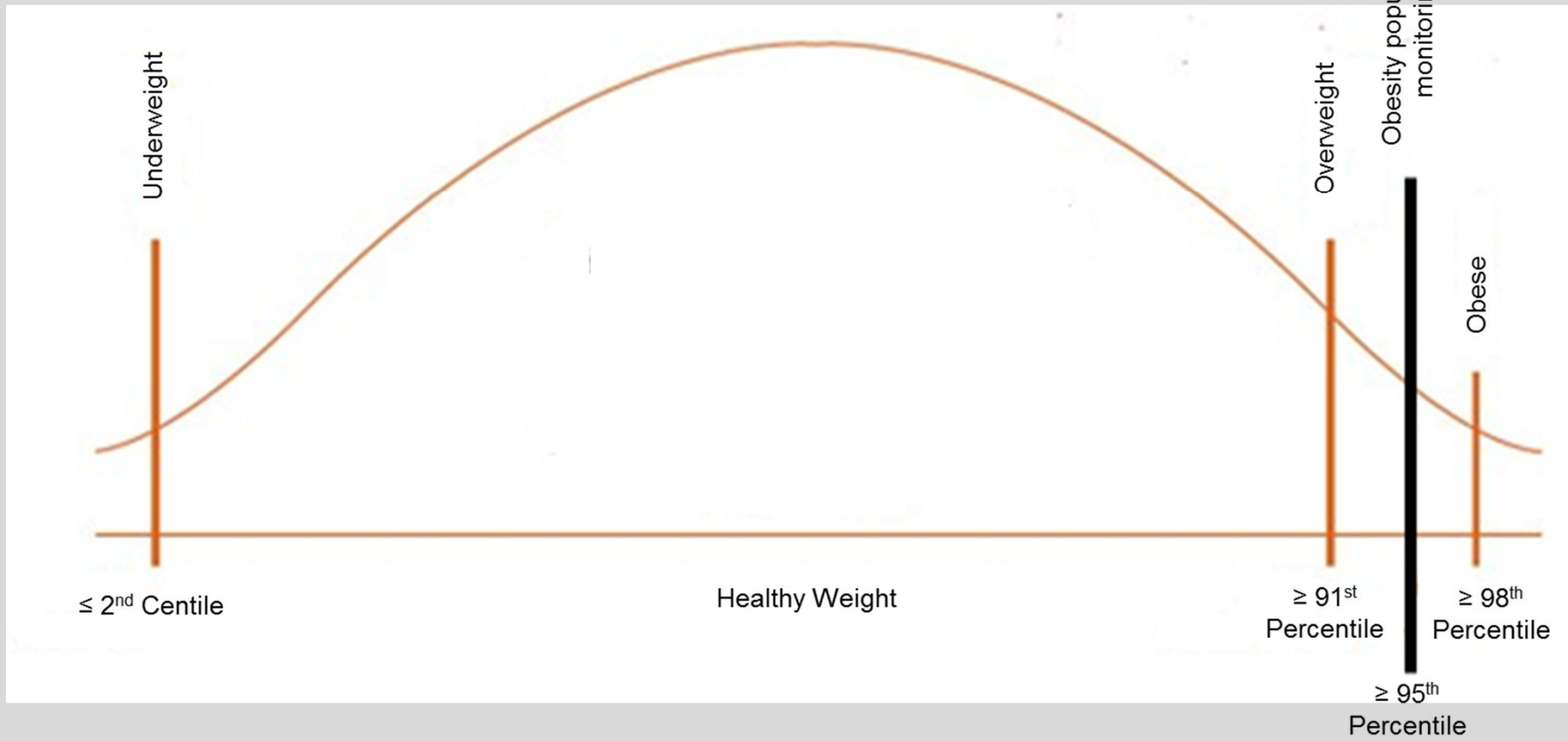
Weight-for-age GIRLS

Birth to 5 years (percentiles)

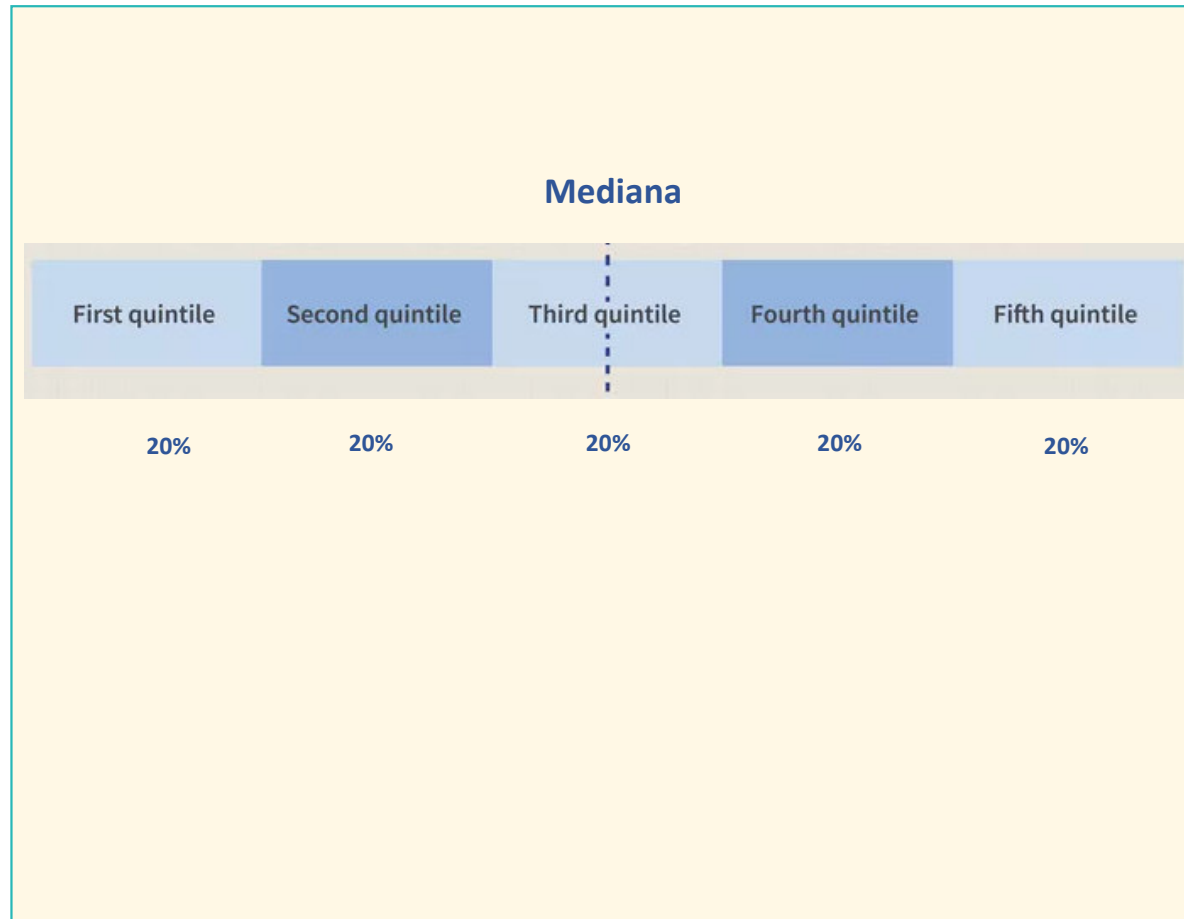


WHO Child Growth Standards

Un neonato alla nascita è sottopeso se ha un peso inferiore a 2500 gr, che corrisponde al 3° percentile.



Quintili



Quintile share ratio

To measure the distribution of income and economic inequality among the participants in a particular economy, economists use different metrics, such as the **income quintile share ratio**.

The **income quintile share ratio** (also called the "S80/S20 ratio") is calculated as the ratio of the total income received by the 20% of the population with the highest income (= 1st quintile) to that income received by the 20% of the population with the lowest (= 5th quintile).

Esempio

	Income (reddito)	Reddito per quintile	% di reddito per quintile
1° quintile	2.000		
	3.000	5.000	7%
	4.000		
	4.000	8.000	12%
	5.000		
5° quintile	6.000	11.000	16%
	8.000		
	9.000	17.000	25%
	12.000		
	15.000	27.000	40%
TOTALE	68.000		

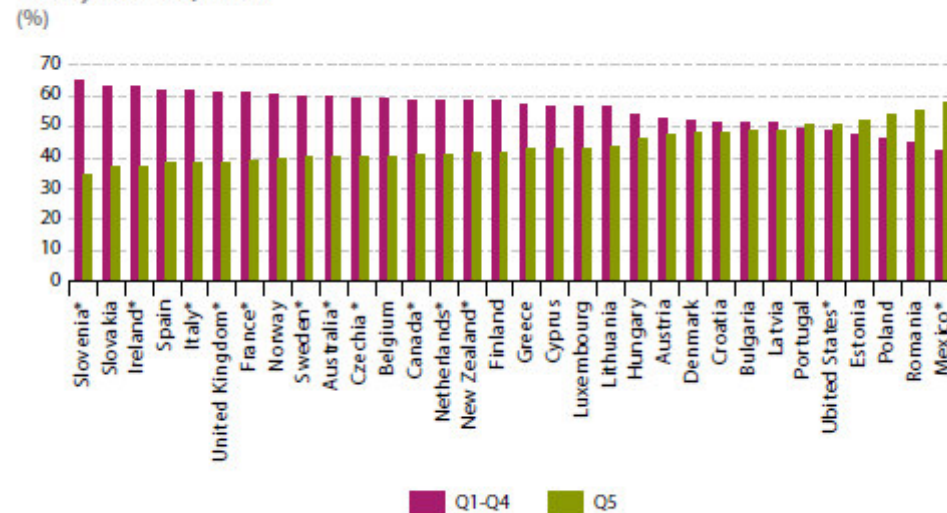
$$\text{Ratio} = \frac{\text{reddito totale del 5° quintile}}{\text{reddito totale del 1° quintile}}$$
$$= 27.000/5.000 = 5,4$$



Perfect equality = 1

Inequality in the distribution of disposable income

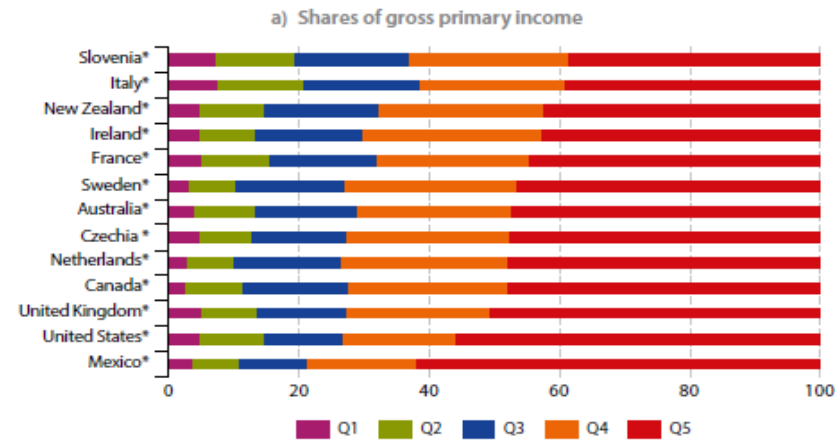
Figure 4.2.1.1: Shares of gross disposable income held by households belonging to the first four quintiles (Q1–Q4), compared with the share held by the fifth quintile (Q5). Year around 2015. Countries ranked by percentage according to the increasing share of disposable income held by the fifth quintile.



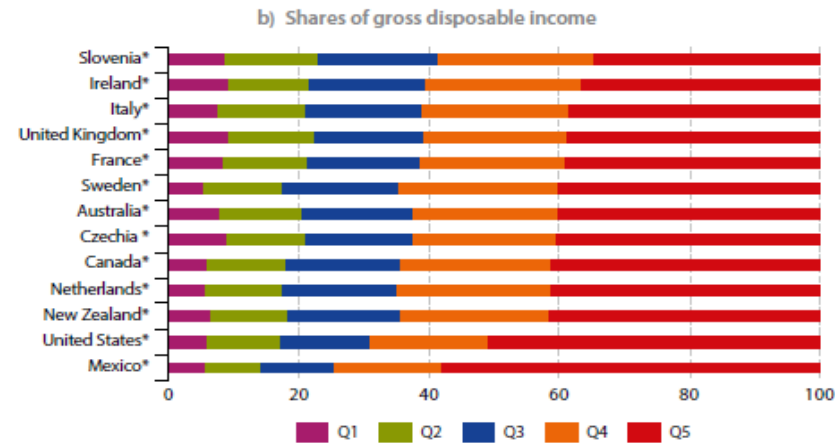
Note: Countries are ranked according to the increasing share of disposable income held by the fifth quintile. Asterisks indicate the countries that carried out a national exercise.

Fonte dei dati: Eurostat

Figure 4.2.2.1: Primary and disposable income shares by equivalised income quantiles.
National exercise countries, year around 2015.
(%)



Note: Countries are ranked according to the decreasing share of gross primary income held by the fifth quintile. Asterisks indicate the countries that have carried out a national exercise.

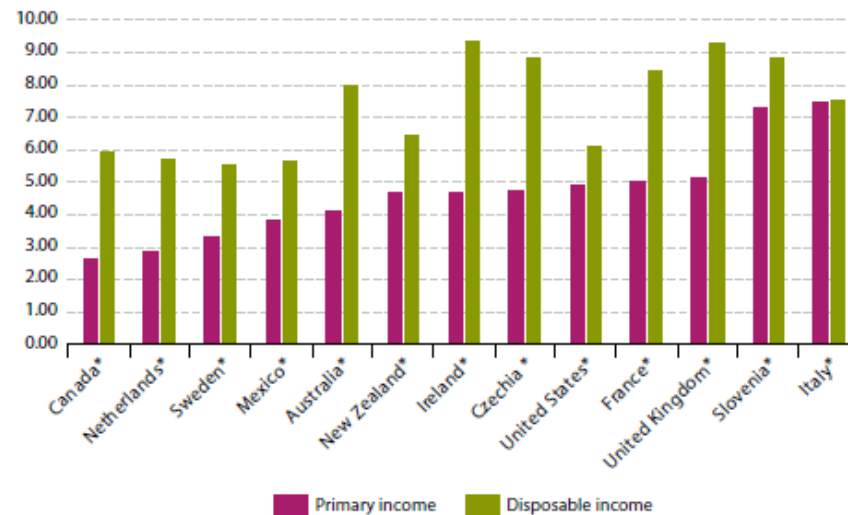


Note: Countries are ranked according to the decreasing share of gross disposable income held by the fifth quintile. Asterisks indicate the countries that have carried out a national exercise.

Fonte dei dati: Eurostat

Fig. 4.2.2.2: Shares of primary and disposable income allocated to the poorest group of households (bottom quintile). National exercise countries, year around 2015.

(%)



Note: Countries are ranked according to the increasing share of primary income held by the first quintile. Asterisks indicate the countries that have carried out a national exercise.

The share of the first income quintile is expected to increase as we move from primary to disposable income. Indeed, this happens for all countries but with significant variations from country to country (see Figure 4.2.2.2). The redistributive effect is greatest in Canada and Ireland, where disposable income doubles as compared with primary income in the lowest income group of households. By contrast, in the United States, the share of disposable income of the same group increases only by 1% as compared with primary income, and in Italy, it does not change at all.

Fonte dei dati: Eurostat

Esercizio

Tabella: Distribuzione del reddito – Anno 2020

Modalità (Reddito USA)	Frequenze assolute (numero di famiglie in mln)
Under \$25.000	23.518
\$25.000-\$50.000	25.596
\$50.000-\$75.000	21.309
\$75.000-\$100.000	15.852
\$100.000-\$125.000	12.360
\$125.000-\$150.000	7.519
\$150.000-\$175.000	6.519
\$175.000-\$200.000	3.875
\$200.000 and over	13.383
Totale	129.931

Fonte dei dati: Census Bureau USA