Probabilità e statistica

- Statistica descrittiva

Insieme di dati $\underline{x} = \{x_1, \dots, x_N\}$, numerosità N.

Frequenza assoluta di un dato

$$f(x_i) = \#\{k = 1, \dots, N : x_k = x_i\}$$

Frequenza relativa di un dato

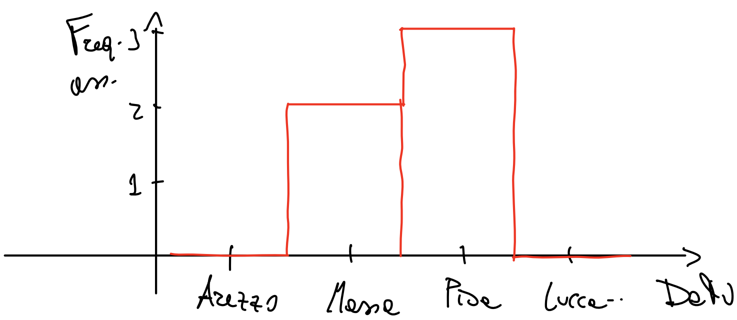
$$p(x_i) = \frac{1}{N} f(x_i)$$

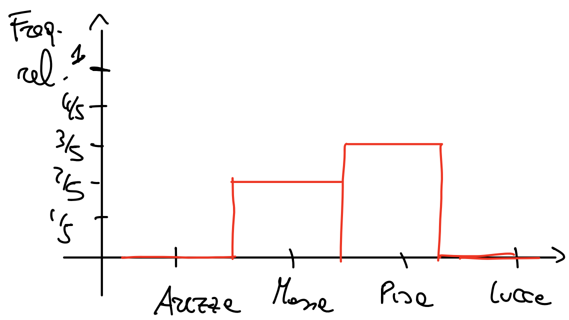
<u>OSS</u> $\quad p(x_i) \geq 0, \quad p(x_i) \leq 1, \quad \sum_{x \text{ dati}} p(x) = 1.$

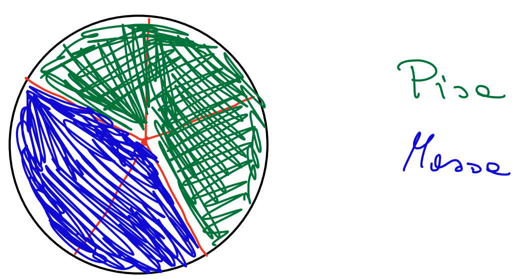
<u>ES</u> $\quad$ Province = { Pisa, Massa, Massa, Pisa, Pisa }, $N = 5$

$$f(\text{Pisa}) = 3, \quad f(\text{Massa}) = 2, \quad \underbrace{f(\text{Livorno}) = f(\text{Firenze}) = \dots = f(\text{Lucca})}_{=\, 0}$$

$$p(\text{Pisa}) = \frac{3}{5}, \quad p(\text{Massa}) = \frac{2}{5}, \quad p(\dots) = 0$$







Supponiamo che $\underline{x} = \{x_1, \dots, x_N\}$ sia numerico, $x_i \in \mathbb{R}$ $\forall i = 1, \dots, N$.

<u>Def</u> Media $\bar{x}$, $\quad \bar{x} := \frac{1}{N} \sum_{i=1}^{N} x_i = \frac{x_1 + x_2 + \dots + x_N}{N}$

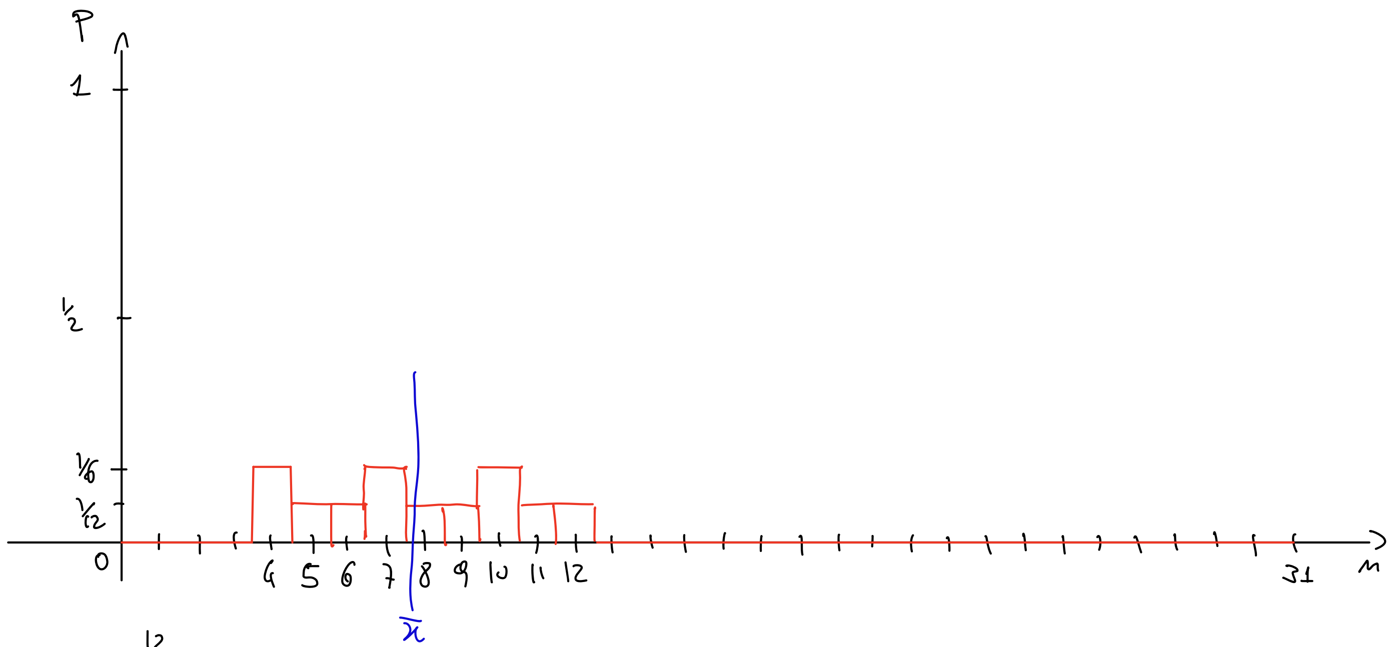
ES $\quad \underline{x} = \{10, 11, 9, 7, 6, 4, 4, 5, 7, 8, 12, 10\}, \quad N = 12$

$f(4) = 2, f(5) = 1, f(6) = 1, f(7) = 2, f(8) = 1, f(9) = 1, f(10) = 2, f(11) = 1, f(12) = 1$

$p(4) = \frac{1}{6}, p(5) = p(6) = \frac{1}{12}, p(7) = \frac{1}{6}, p(8) = p(9) = \frac{1}{12}, p(10) = \frac{1}{6}, p(11) = p(12) = \frac{1}{12}$

$f(n) = 0, \quad p(n) = 0 \quad \forall n \notin \{4, 5, 6, 7, 8, 9, 10, 11, 12\}$

$n \in \mathbb{N}_0, \quad n \geq 0, \quad n \leq 31.$



$$\bar{x} = \frac{1}{12} \sum_{i=1}^{12} x_i = \frac{1}{12}(10+11+9+7+6+4+4+5+7+8+12+10) = \frac{93}{12} = 7{,}75$$

$\underline{x} = \{10, 11, 9, 7, 6, 4, 4, 5, 7, 8, 12, 10\}$

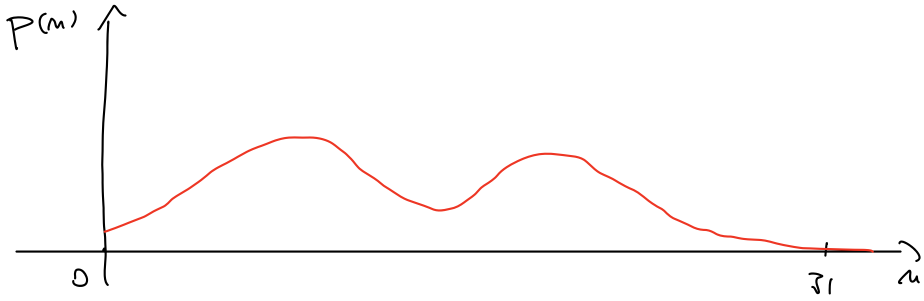
OSS $\quad \bar{x}$ è il punto in cui si realizza $\min\limits_{a \in \mathbb{R}} \sum\limits_{i=1}^{n} (x_i - a)^2$

OSS $\quad \frac{1}{N}\sum\limits_{i=1}^{N} x_i = \sum\limits_{i=1}^{N} \frac{1}{N} x_i = \sum\limits_{n=0}^{31} \frac{1}{N} f(n)\, n = \sum\limits_{n=0}^{31} p(n)\, n$

$$\bar{x} = \frac{2}{12}\cdot 4 + \frac{1}{12}\cdot 5 + \frac{1}{12}\cdot 6 + \frac{2}{12}\cdot 7 + \frac{1}{12}\cdot 8 + \frac{1}{12} 9 + \frac{2}{12}\cdot 10 + \frac{1}{12}\cdot 11 + \frac{1}{12}\cdot 12$$

$f(4) = 2, f(5) = 1, f(6) = 1, f(7) = 2, f(8) = 1, f(9) = 1, f(10) = 2, f(11) = 1, f(12) = 1$

$p(4) = \frac{1}{6}, p(5) = p(6) = \frac{1}{12}, p(7) = \frac{1}{6}, p(8) = p(9) = \frac{1}{12}, p(10) = \frac{1}{6}, p(11) = p(12) = \frac{1}{12}$



Def Varianza (empirica) $\mathrm{var}(\underline{x})$, Scarto quadratico medio (empirico) o deviazione standard (empirica) $\sigma(\underline{x})$.

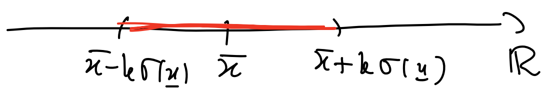
$$\mathrm{var}(\underline{x}) := \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2 \quad , \quad \sigma(\underline{x}) := \sqrt{\mathrm{var}(\underline{x})}$$

Disuguaglianza di Chebyshev

Sia $d > 0$, allora
$$\frac{\#\{x_i : |x_i - \bar{x}| > d\}}{N} \leq \frac{\mathrm{var}(\underline{x})}{d^2}$$

ES $d = k\,\sigma(\underline{x})$, $k \in \mathbb{N}$

$$|x_i - \bar{x}| > k\sigma(\underline{x}) \iff x_i \notin (\bar{x} - k\,\sigma(\underline{x}), \bar{x} + k\,\sigma(\underline{x}))$$



$$\frac{\#\{x_i : |x_i - \bar{x}| > k\,\sigma(\underline{x})\}}{N} \leq \frac{\mathrm{var}(\underline{x})}{k^2 \sigma^2(\underline{x})} = \frac{1}{k^2}$$

OSS
$$\frac{\#\{x_i : |x_i - \bar{x}| \leq k\,\sigma(\underline{x})\}}{N} \geq 1 - \frac{1}{k^2}$$

OSS
$$\mathrm{var}(\underline{x}) = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i^2 - 2x_i\bar{x} + \bar{x}^2) =$$
$$= \frac{1}{N} \sum_{i=1}^{N} x_i^2 - \frac{2}{N} \sum_{i=1}^{N} x_i \bar{x} + \frac{1}{N} \sum_{i=1}^{N} \bar{x}^2 =$$

$$= \frac{1}{N}\sum_{i=1}^{N} x_i^2 - 2\bar{x}\cdot \underbrace{\frac{1}{N}\sum_{i=1}^{N} x_i}_{\bar{x}} + \bar{x}^2 \cdot \underbrace{\frac{1}{N}\sum_{i=1}^{N} 1}_{1}$$

$$= \frac{1}{N}\sum_{i=1}^{N} x_i^2 - 2\bar{x}^2 + \bar{x}^2$$

$$\boxed{\operatorname{var}(\underline{x}) = \frac{1}{N}\sum_{i=1}^{N} x_i^2 - \bar{x}^2}$$

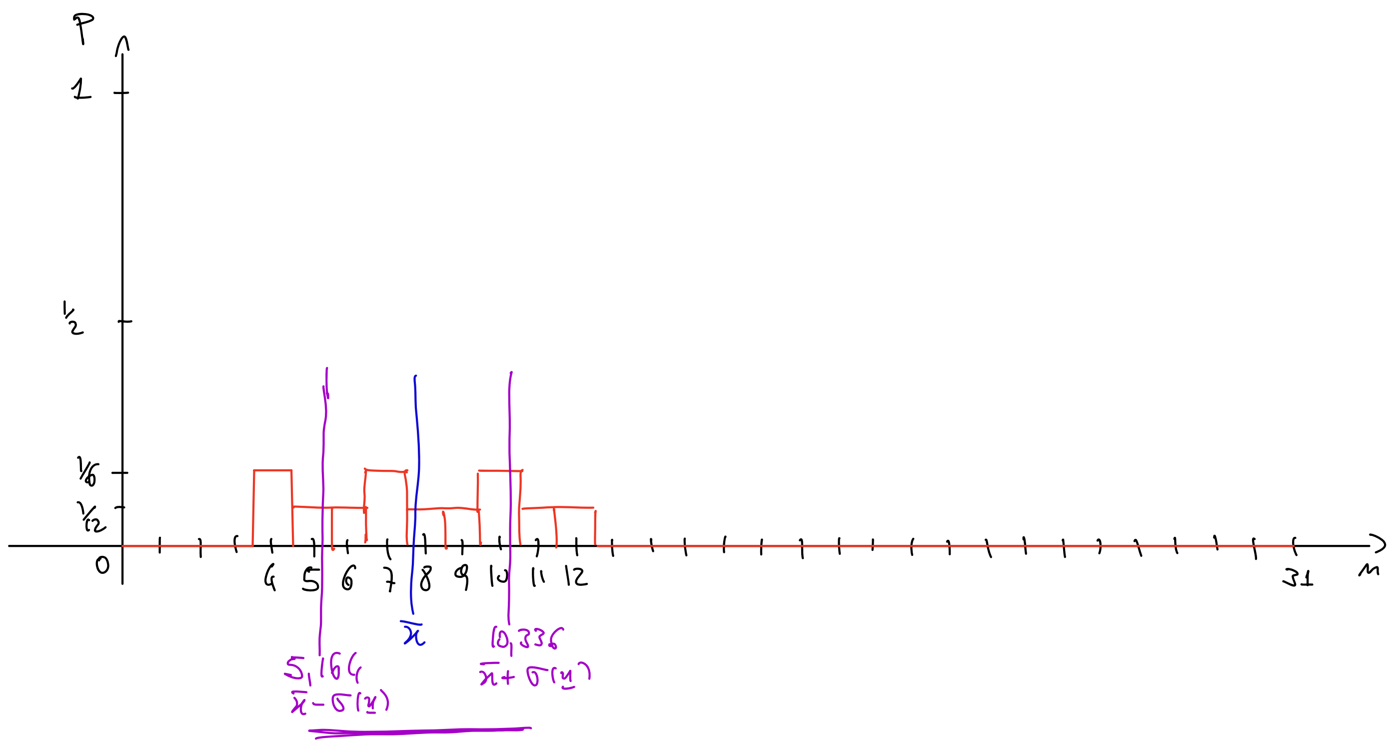
ES $\quad \frac{1}{N}\sum_{i=1}^{N} x_i^2 - \bar{x}^2 = \sum_{n=0}^{31} p(n)\, n^2 - \bar{x}^2$

$$\operatorname{var}(\underline{x}) = \left(\frac{1}{6}\cdot 4^2 + \frac{1}{12}\cdot 5^2 + \frac{1}{12}\cdot 6^2 + \frac{1}{6}\cdot 7^2 + \frac{1}{12}\, 8^2 + \frac{1}{12}\, 9^2 + \frac{1}{6}\, 10^2 + \frac{1}{12}\cdot 11^2 + \frac{1}{12}\cdot 12^2\right) - (7{,}75)^2$$

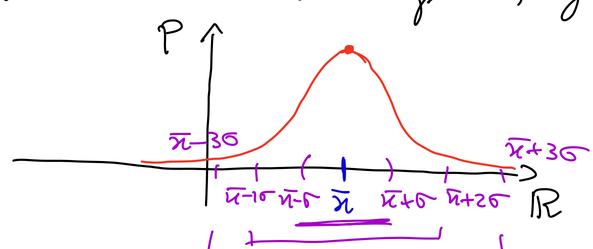
$$= 66{,}75 - 60{,}0625 = 6{,}6875 \ d^2$$

$p(4) = \frac{1}{6}$, $p(5) = p(6) = \frac{1}{12}$, $p(7) = \frac{1}{6}$, $p(8) = p(9) = \frac{1}{12}$, $p(10) = \frac{1}{6}$, $p(11) = p(12) = \frac{1}{12}$

$$\sigma(\underline{x}) = \sqrt{6{,}6875} \sim 2{,}586 \ d$$



Esempi $\quad$ Distribuzione normale (a campana, gaussiana)



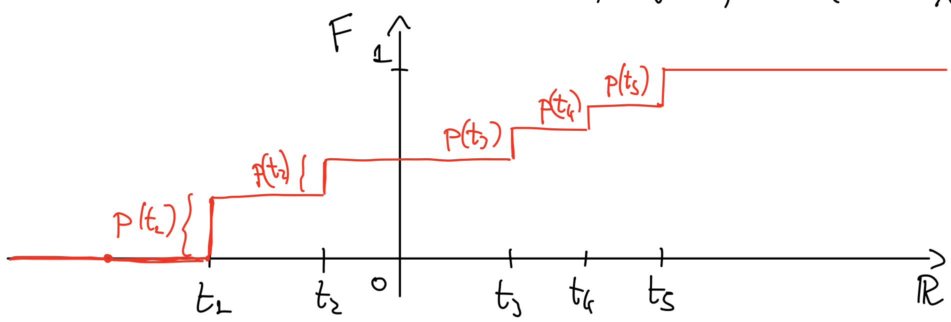
$$\frac{\#\{x_i : |x_i - \bar{x}| \le \sigma(x)\}}{N} \ge \frac{68}{100}$$

$$\frac{\#\{x_i : |x_i - \bar{x}| \le 2\sigma(x)\}}{N} \ge \frac{95}{100}$$

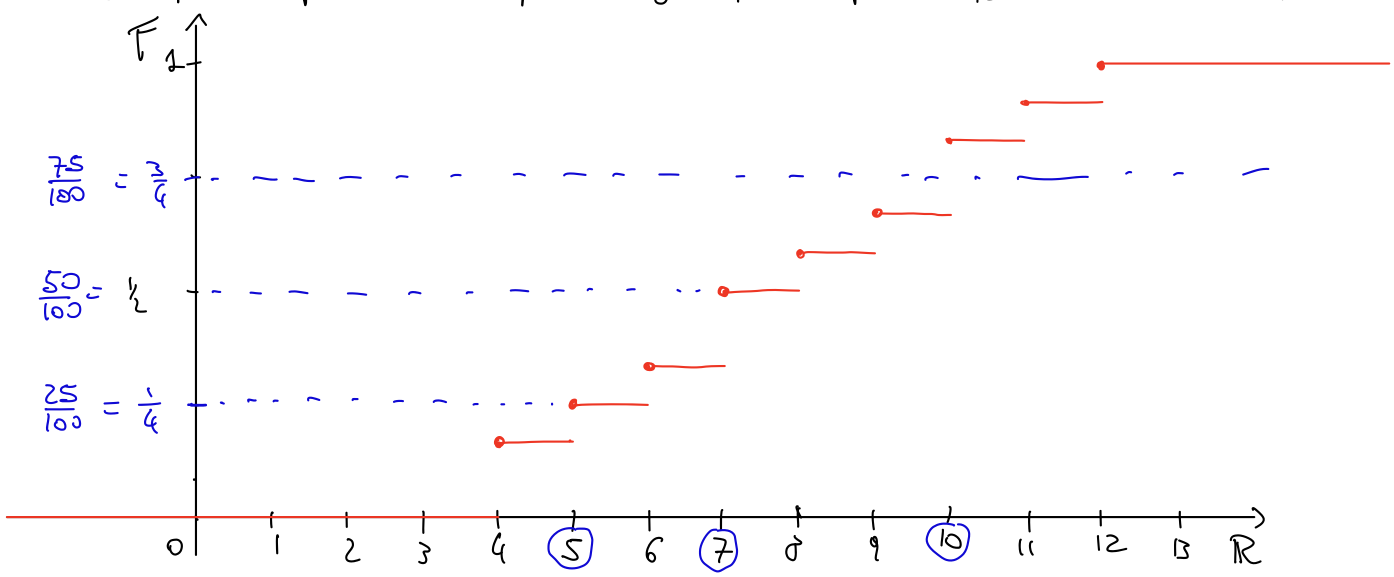
$$\frac{\#\{x_i : |x_i - \bar{x}| \le 3\sigma(x)\}}{N} \ge \frac{99{,}7}{100}$$

Def Funzione di ripartizione empirica di $x = \{x_1, \dots, x_N\}$, N numerosità

$$F : \mathbb{R} \to [0,1] \quad , \quad F(t) := \frac{1}{N} \#\{x_i : x_i \le t\}$$



$p(4) = \frac{1}{6}$, $p(5) = p(6) = \frac{1}{12}$, $p(7) = \frac{1}{6}$, $p(8) = p(9) = \frac{1}{12}$, $p(10) = \frac{1}{6}$, $p(11) = p(12) = \frac{1}{12}$



Def Sia $0 < k < 100$, si chiama k-simo percentile il numero t_k tale che ci siano almeno $\frac{k}{100}$ di dati $\le t_k$ e almeno $1 - \frac{k}{100}$ di dati $\ge t_k$.

25-simo percentile o quartile, $k = 25$, 0,25 quantile

50-esimo percentile o mediana, $k=50$, 0,5 quantile

75-esimo percentile o terzo quartile, $k=75$. , 0,75 quantile

β-quantile con $\beta = \frac{k}{100}$ $\longleftrightarrow$ k-esimo percentile