Ex. 5 (Scheda 6)

$D =$ " risultato del lancio di un dado a tre facce "

Si lanciano D monete:

$T =$ " numero di teste "

a) Congiunta e marginali di D e T

b) $\mathbb{E}[T] = ?$

a)

D	1	2	3
p_D	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

$$S_T = \{0, 1, 2, 3\}$$

$D \setminus T$	0	1	2	3	p_D
1	$\frac{1}{2}\cdot\frac{1}{3}$	$\frac{1}{2}\cdot\frac{1}{3}$	0	0	$1/3$
2	$\frac{1}{4}\cdot\frac{1}{3}$	$\frac{1}{2}\cdot\frac{1}{3}$	$\frac{1}{4}\cdot\frac{1}{3}$	0	$1/3$
3	$\frac{1}{8}\cdot\frac{1}{3}$	$\frac{3}{8}\cdot\frac{1}{3}$	$\frac{3}{8}\cdot\frac{1}{3}$	$\frac{1}{8}\cdot\frac{1}{3}$	$1/3$
p_T	$\frac{7}{24}$	$\frac{11}{24}$	$\frac{5}{24}$	$\frac{1}{24}$	1

b) $\mathbb{E}[T] = \sum_{i=0}^{3} i\, p_T(i)$

$= 1$

$$p_{(D,T)}(1,0) = \mathbb{P}(D=1, T=0) \underset{\substack{\uparrow \\ \text{Reg.}}}{=} \mathbb{P}(T=0 \mid D=1)\,\mathbb{P}(D=1) = \frac{1}{2}\,\frac{1}{3}$$

catena

$$p_{(D,T)}(2,1) = \mathbb{P}(D=2, T=1) = \mathbb{P}(T=1 \mid D=2)\,\mathbb{P}(D=2) = \frac{1}{2}\,\frac{1}{3}$$

Legge condizionata di Y a un evento B

PROB. CONDIZIONATA: $\mathbb{P}(A|B) = \dfrac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}$ $(\mathbb{P}(B)>0)$

Y v.a. DISCRETA:

$$p_Y(y_j|B) = \frac{\mathbb{P}((Y=y_j)\cap B)}{\mathbb{P}(B)}$$

Ad esempio: $B=(X=x_i)$, X v.a. DISCRETA

$$p_Y(y_j|x_i) = \frac{\mathbb{P}(Y=y_j, X=x_i)}{\mathbb{P}(X=x_i)}$$

cioè

$$p_Y(y_j|x_i) = \frac{p_{(X,Y)}(x_i, y_j)}{p_X(x_i)}$$

Ex. 5 (Scheda 6)

$T \mid D=1$	0	1
$p_T(\cdot \mid 1)$	$\frac{1}{2}$	$\frac{1}{2}$

$$T \mid D=1 \sim B\left(\frac{1}{2}\right) = B\left(1, \frac{1}{2}\right)$$

$T \mid D=2$	0	1	2
$p_T(\cdot \mid 2)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

$$T \mid D=2 \sim B\left(2, \frac{1}{2}\right)$$

$$T \mid D=3 \sim B\left(3, \frac{1}{2}\right)$$

STATISTICA

1) CAMPIONAMENTO STATISTICO

2) STATISTICA DESCRITTIVA

3) STATISTICA INFERENZIALE / MATEMATICA

POPOLAZIONE / CAMPIONE

↑ TUTTI I POSSIBILI LANCI DI UN DADO

ogni elemento della popolazione si chiama UNITÀ STATISTICA

Caratteristiche di una UNITÀ STATISTICA: VARIABILI, X, Y, Z, ...

CAMPIONAMENTO STATISTICO: trarre un campione il più rappresentativo possibile

STATISTICA DESCRITTIVA
- → VAR. DISCRETE
- → VAR. CONTINUE
- → VAR. QUALITATIVE

VAR. DISCRETA: VOTI FINALI IN INGLESE

$x_1 = 7$; $x_2 = 6$; $x_3 = 8$, ...

VAR. CONTINUA: I risultati (in metri) di salto in lungo da fermo:

$x_1 = 1.36$; $x_2 = 1.46$; $x_3 = 1.56$ cm, ...

VAR. QUALITATIVE: Questionario in cui viene chiesto di indicare l'attività preferita:

A = amici, S = sport, ..., N = altre attività

$x_1 = A$, $x_2 = S$, $x_3 = A$, $x_4 = N$

FREQUENZE

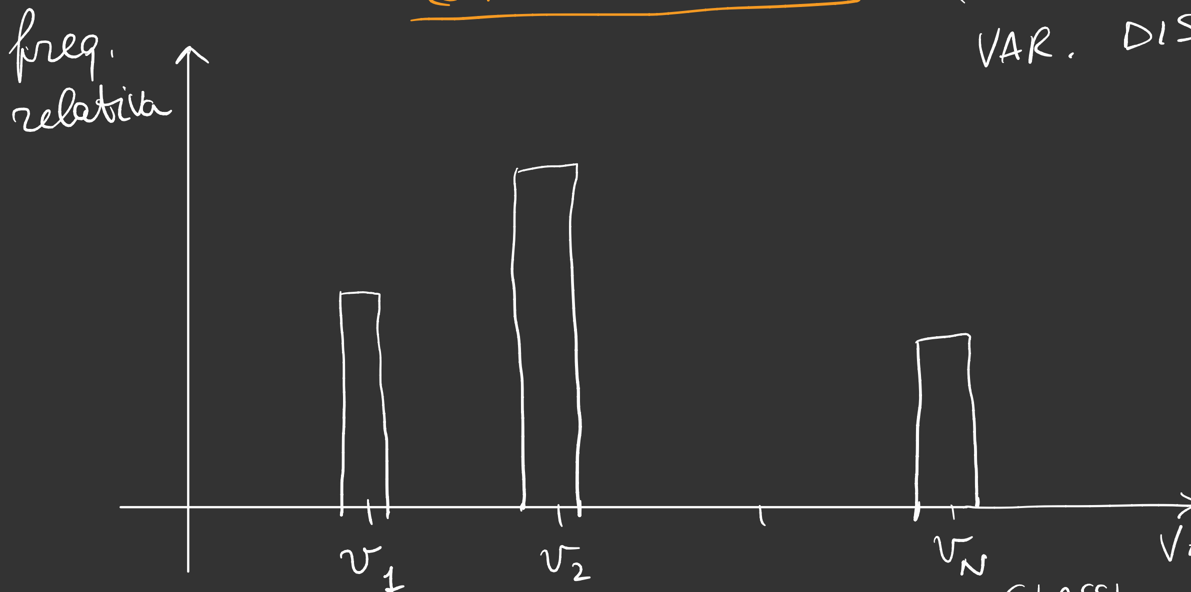
X variabile (qual. o quant.) che assume i valori: $v_1, v_2, \dots, v_N$

VALORE	FREQUENZA ASSOLUTA	FREQ. RELATIVA
v_1	n_1	$f_1 = \frac{n_1}{n}$
v_2	n_2	$f_2 = \frac{n_2}{n}$
$\vdots$		
v_N	n_N	$f_N = \frac{n_N}{n}$

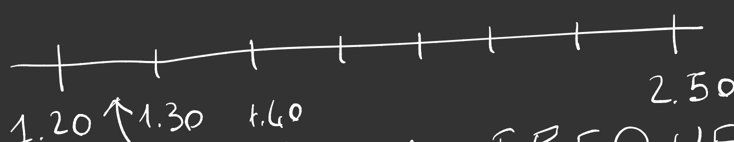
ESEMPIO delle ATTIVITÀ: $v_1 = A$, $v_2 = S$, ...

ORTOGRAMMA

(VAR. QUALITATIVE
VAR. DISCRETE)



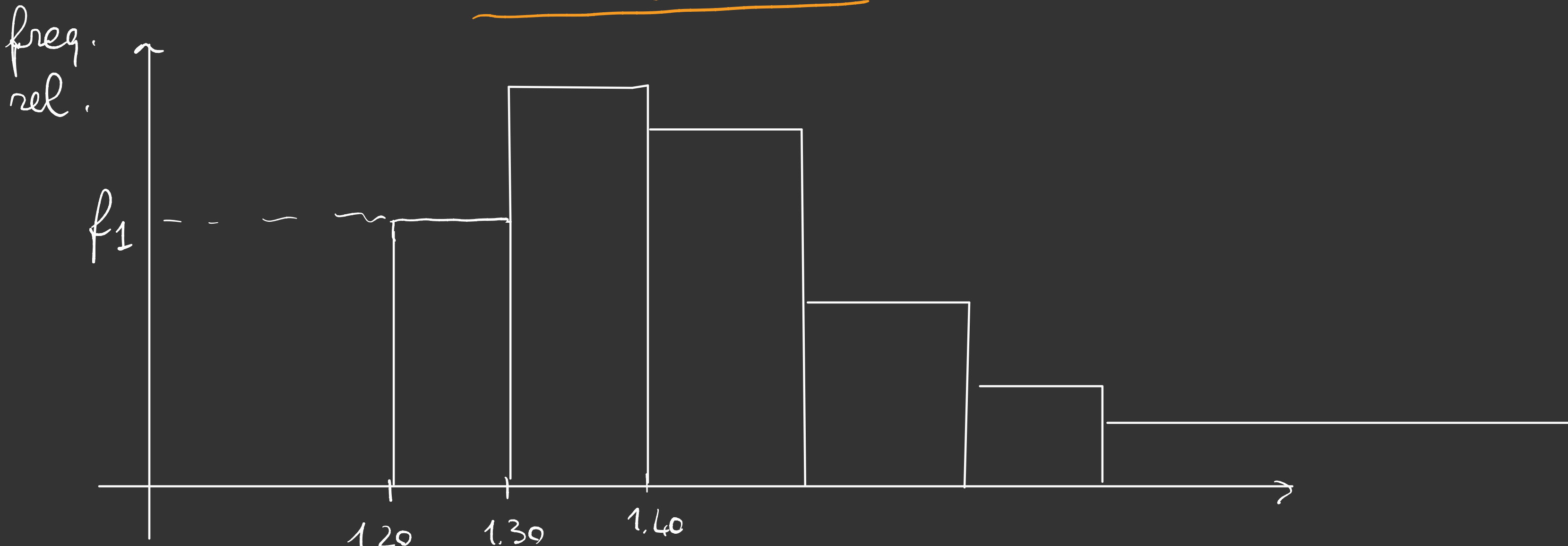
VAR. CONTINUE



CLASSI DI FREQUENZA

CLASSI di FREQUENZA	freq. ass.
1,20 – 1.30	n_1
1.30 – 1.40	n_2

ISTOGRAMMA



INDICI DI SINTESI

OUTLIER

INDICI DI POSIZIONE:

MEDIA: $\bar{x}_n = \dfrac{x_1 + \dots + x_n}{n} = \dfrac{\sum_{i=1}^{N} v_i n_i}{n} = \sum_{i=1}^{N} v_i f_i$

MEDIANA: dati in ordine crescente:

$x_{(1)}, x_{(2)}, \dots, x_{(n)}$

$$Me_n = \begin{cases} \dfrac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2} & n \text{ PARI} \\ x_{([\frac{n}{2}]+1)} & n \text{ DISPARI} \end{cases}$$

MODA: $Mo = v_i$ più frequente

INDICI DI DISPERSIONE

VARIANZA (dev. standard)

Scarti dalla media: $x_i - \bar{x}$

Scarti assoluti: $|x_i - \bar{x}|$

Scarti quadratici: $(x_i - \bar{x})^2$

Varianza campionaria:

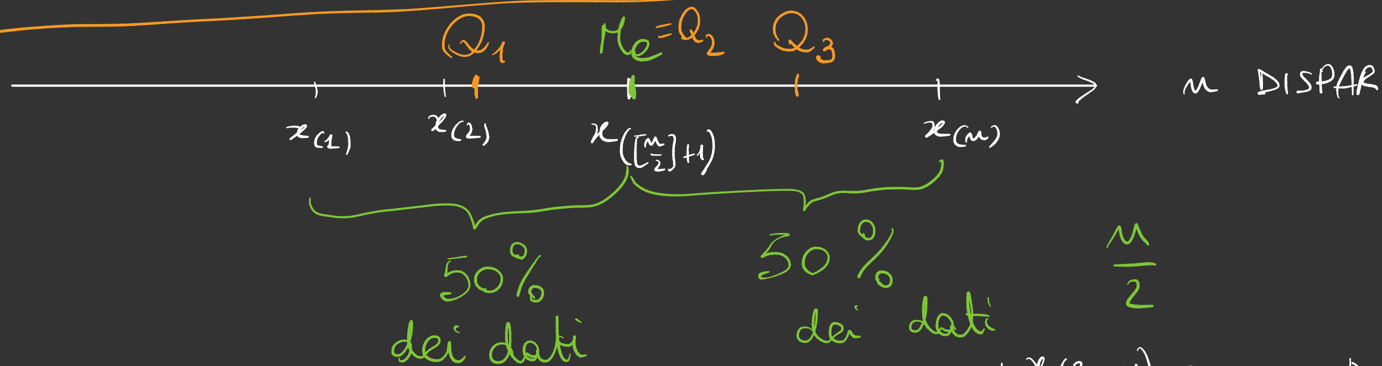
$$S_n^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$S_{Me} = \frac{1}{n} \sum_{i=1}^{n} |x_i - Me|$$

Scostamento semplice medio

RANGE: $R = x_{(n)} - x_{(1)}$

DIFFERENZA INTERQUARTILE

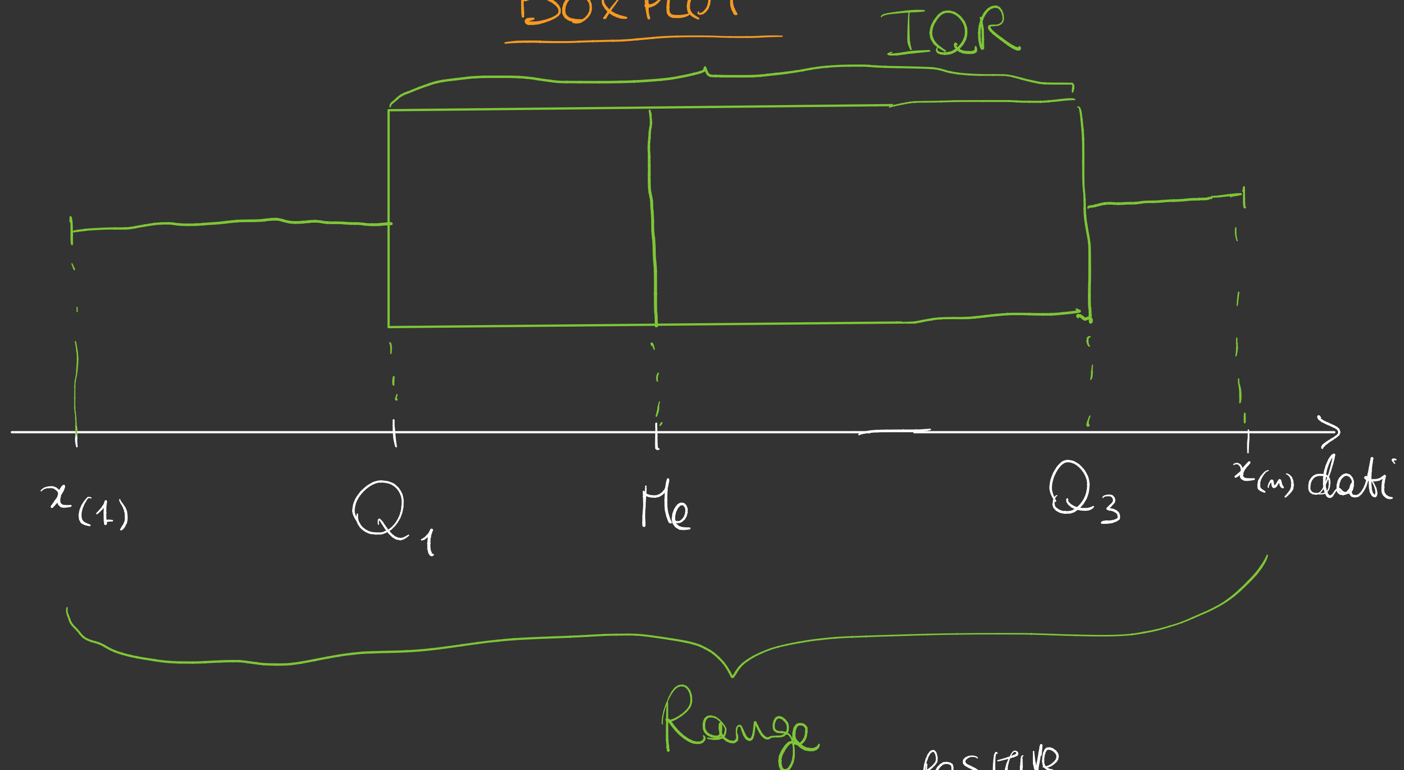


$$Q_1 = \begin{cases} \dfrac{x_{(\frac{n}{4})} + x_{(\frac{n}{4}+1)}}{2}, & \dfrac{n}{4} \in \mathbb{N} \\ x_{([\frac{n}{4}]+1)}, & \dfrac{n}{4} \notin \mathbb{N} \end{cases}$$

$$Q_3 = \begin{cases} \dfrac{x_{(\frac{3}{4}n)} + x_{(\frac{3}{4}n+1)}}{2}, & \dfrac{3}{4}n \in \mathbb{N} \\ x_{([\frac{3}{4}n]+1)}, & \dfrac{3}{4}n \notin \mathbb{N} \end{cases}$$

$$IQR = Q_3 - Q_1$$

BOX PLOT



INDICI DI ASIMMETRIA:

$$\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^3$$

POSITIVO

NEGATIVO

Teoremi Limite

f_j^n = freq. relativa del valore v_j con riferimento al dataset $x_1, \dots, x_n$

$$\bar{x}_n = \frac{x_1 + \dots + x_n}{n} = \sum_{j=1}^{N} v_j f_j^n$$

$$\bar{x}_n \xrightarrow[n \to +\infty]{} \mathbb{E}[X]$$

$$f_j^n \xrightarrow[n \to +\infty]{} \mathbb{P}(X = v_j) = p_X(v_j)$$

v.a. X = "risultato del lancio di un dado"

Lancio il dado n volte:

$x_1 = 5$; $x_2 = 3$; $x_3 = 4$; ...

$X_1, \quad X_2, \quad \dots, \quad X_n, \dots$

$(X_n)_{n \in \mathbb{N}}$

Definizione $(X_n)_{n \in \mathbb{N}}$ è una successione di v.a. i.i.d. (indipendenti e identicamente distribuite):

1) $X_1, X_2, \dots, X_n, \dots$ hanno la stessa distribuzione

2) $X_1, X_2, \dots, X_n, \dots$ sono indipendenti

LGN: Legge dei grandi numeri

Teorema

Sia $(X_n)_{n\in\mathbb{N}}$ una successione di v.a. i.i.d. con media μ e varianza σ^2. Allora, posto

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n} \quad \text{(media campionaria)}$$

si ha:

$$\forall \varepsilon > 0, \quad \lim_{n\to+\infty} \mathbb{P}(|\bar{X}_n - \mu| > \varepsilon) = 0.$$

$$f_j^n \longrightarrow \mathbb{P}(X = v_j)$$

$$\left(\bar{x}_n \longrightarrow \mathbb{E}[X] = \mu \right)$$

$$Y_j = \begin{cases} 1, & \text{se } X = v_j \\ 0, & \text{se } X \neq v_j \end{cases}$$

$$\mathbb{E}[Y_j] = \mathbb{P}(X = v_j)$$

$A = (X = v_j)$

$$Y = 1_A \implies \mathbb{E}[Y] = \mathbb{P}(A)$$

$$\bar{y}_n = \frac{y_1 + \dots + y_n}{n} = \frac{\#\text{ che si verifica } A}{n} = f^n(A)$$

METODO MONTE CARLO

VOGLIO APPROX $\int_0^1 f(x)\,dx = \mathbb{E}[f(U)]$, $U \sim Unif(0,1)$

$X = f(U)$

numeri pseudocasuali: $u_1, u_2, \dots, u_n, \dots$

$x_1 = f(u_1),\ x_2 = f(u_2), \dots, x_n = f(u_n), \dots$

$$\bar{x}_n = \sum_{i=1}^{n} \frac{x_i}{n} = \sum_{i=1}^{n} \frac{f(u_i)}{n} \overset{LGN}{\approx} \int_0^1 f(x)\,dx$$

$$\frac{1}{\sqrt{n}}$$