

Esercizi di Statistica per Biotecnologie

Francesca Pizzorni Ferrarese

Parte I – Statistica descrittiva

Es.1

Rilevando con uno strumento di misurazione il numero di particelle cosmiche in 40 periodi consecutivi di un minuto si ottengono i seguenti dati

0	2	1	4	3	1	2	3	8	2	5	2	1	3	3	1	3	2	2	5
4	4	4	2	3	5	5	1	1	2	4	4	2	3	3	3	3	3	3	2

1. Indicare il tipo di carattere
2. Riportare in una tabella ad entrata semplice le frequenze relative, assoluta e cumulativa (tabella di distribuzione di frequenza).
3. Scegliere una rappresentazione grafica idonea

Es.2

I seguenti dati sono il risultato di 80 determinazioni, in una data unità di misura, dell'emissione giornaliera di gas inquinante da un impianto industriale

15.8	26.4	17.3	11.2	23.9	24.8	18.7	13.9	9.0	13.2
22.7	9.8	6.2	14.7	17.5	26.1	12.8	28.6	17.6	23.7
26.8	22.7	18.0	20.5	11.0	20.9	15.5	19.4	16.7	10.7
19.1	15.2	22.9	26.6	20.4	21.4	19.2	21.6	16.9	19.0
18.5	23.0	24.6	20.1	16.2	18.0	7.7	13.5	23.5	14.5
14.4	29.6	19.4	17.0	20.8	24.3	22.5	24.6	18.4	18.1
8.3	21.9	12.3	22.3	13.3	11.8	19.3	20.0	25.7	31.8
25.9	10.5	15.9	27.5	18.1	17.9	9.4	24.1	20.1	28.5

1. Indicare il tipo di carattere
2. Calcolare il campo di variazione o range dei dati
3. Riportare in una tabella ad entrata semplice le frequenze relative, assoluta e cumulative (tabella di distribuzione di frequenza) per ciascuna delle classi opportunamente identificate.
4. Rappresentare mediante istogramma i dati

Es.3

In uno stabilimento vengono registrati i casi di malfunzionamento di una macchina utensile controllata dal computer, e le loro cause. I dati relativi ad un certo mese sono i seguenti

fluttuazioni di tensione	6
instabilità del sistema di controllo	22
errore dell'operatore	13
strumento usurato e non sostituito	2
altre cause	5
<i>Totale</i>	48

1. Indicare il tipo di carattere
2. Riportare in una tabella ad entrata semplice le frequenze relative e assoluta (tabella di distribuzione di frequenza) per ciascuna delle classi opportunamente identificate.

Es. 4

Sia dato il seguente insieme di 20 dati, che rappresentano il peso alla nascita (in g) di 20 bambini nati in una settimana in clinica

3280	3320	2500	2760
3260	3650	2840	3250
3240	3200	3600	3320
3480	3020	2840	3200
4160	2580	3540	3780

1. Calcolare la media e la mediana
2. Calcolare la media e la mediana sostituendo in tabella al primo valore 500 (nato prematuro). Commentare i risultati.

Es. 5

I dati seguenti rappresentano i valori dei globuli bianchi (in migliaia) rilevati in 10 pazienti ricoverati in una mattina in ospedale

7 35 5 9 8 3 10 12 8 7

1. Calcolare la media e la mediana
2. Calcolare la media e la mediana nel caso il secondo paziente in tabella avesse un valore di 70 anziché 35. Commentare i risultati.

Es. 6

Calcolare il primo ed il terzo quartile dell'insieme di dati

32.2 32.0 30.4 31.0 31.2 31.3 30.3 29.6 30.5 30.7

Es.7

Calcolare la varianza e la deviazione standard dei dati presentati nell'Es. 4

Es. 8

I seguenti dati sono i tempi di esecuzione di una certa operazione misurati in minuti

0.6 1.2 0.9 1.0 0.6 0.8

Calcolare la varianza e la deviazione standard

Es. 9

Per la partecipazione ad una gara di matematica una scuola deve formare una squadra da 6 studenti; con una selezione preliminare, attraverso un test con punteggio Massimo di 100 punti, sulla base della media dei migliori 6 punteggi risultano 3 squadre a pari merito. Con quale criterio può essere scelta la squadra da mandare alla gara?

<i>squadra</i>	<i>punteggi degli studenti</i>					
A	73	76	77	85	88	90
B	74	74	78	84	88	91
C	72	77	79	82	84	95

Es. 10

Le misure del diametro di un cuscinetto a sfera effettuate con uno strumento hanno valor medio $\bar{x}=3.92$ mm e uno scarto quadratico medio $\sigma=0.015$ mm; le misure della lunghezza di una sbarra rigida effettuate con un altro strumento hanno invece un valore medio $\bar{x}=1.54$ m e uno scarto quadratico medio $\sigma=0.008$ m. Quale dei due strumenti è relativamente più preciso?

Es. 11

Nella tabella seguente si riportano i punteggi ottenuti in 40 lanci successive di un dado

<i>classe (punteggio)</i>	<i>f_i</i>
1	9
2	8
3	5
4	5
5	6
6	7

Calcolare la media, la mediana, la moda e la varianza.

Es. 12

In un collettivo di 9 imprese è stato rilevato il numero di addetti:

Impresa	1	2	3	4	5	6	7	8	9
Numero addetti	300	50	200	50	150	80	150	100	150

1. Determinare media e varianza.
2. Costruire il box-plot della distribuzione.

Es. 13

Sia data la seguente distribuzione di frequenza di 283 dipendenti di un'azienda per classi di età:

classi di età	30 - 34	34 - 38	38 - 42	42 - 46	46 - 50	50 -52	52 -56	56 -58	58 -65
dipendenti	37	48	20	33	33	21	26	50	15

1. Si costruisca istogramma e box-plot della distribuzione.
2. Si commenti la natura simmetrica o asimmetrica della distribuzione.

Es. 14

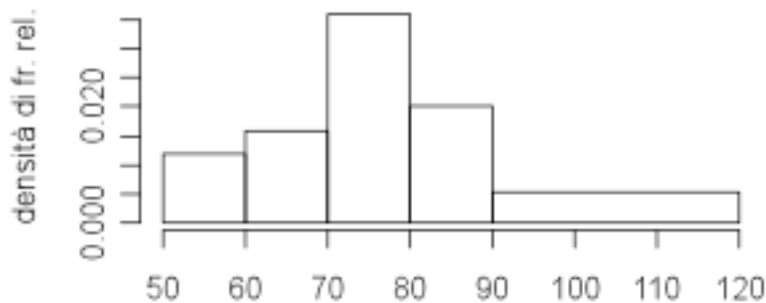
Graficare mediante box-plot i seguenti dati: età delle attrice all'epoca in cui hanno vinto il primo oscar

50	44	35	80	26	28	41
21	61	38	49	33	74	30
33	41	31	35	41	42	37
26	34	34	35	26	61	60
34	24	30	37	31	27	39
34	26	25	33			

Indicare la presenza di eventuali outlier.

Es. 15

L'istogramma rappresenta l'incidenza di una patologia per fascia d'età registrata nel 2009 in un ospedale (N=100).



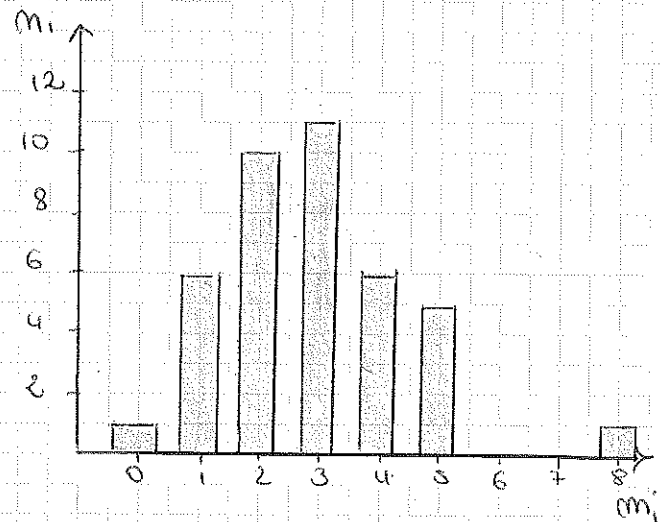
- 1) quanti elementi si ammalano prima dei 70 anni?
- 2) quanti elementi si ammalano fra i 70 e i 90 anni?
- 3) Indicare la media
- 4) Indicare la moda
- 5) Indicare la mediana.
- 6) Indicare la distanza interquartile.

STATISTICA DESCRITTIVA

① 1. quantitativo discreto.

l_i	m_i	f_i	F_i
0	1	0,025	0,025
1	6	0,15	0,175
2	10	0,25	0,425
3	11	0,275	0,7
4	6	0,15	0,85
5	5	0,125	0,975
6	0	0	0,975
7	0	0	0,975
8	1	0,025	1

3. diagramma a barre



⑤ Ordino i dati in maniera crescente.

1. 3 5 7 7 8 8 9 10 12 35

$$\bar{m} = \frac{3+5+\dots+12+35}{10} = 10,4 \quad H = 8$$

$$2. \bar{m} = \frac{3+5+\dots+12+35}{10} = 13,9 \quad H = 8 \begin{matrix} \uparrow \\ 70 \end{matrix}$$

La media (risultato) meno sensibile agli outlier rispetto alla media.

⑥ Ordino i dati.

29,6 30,3 30,4 30,5 30,7 31,0 31,2 31,3
 32,0 32,2

Calcolo la dimensione dei "quartili" $\frac{N-1}{4} = 2,25 \rightarrow 3$

Se mi trovo oggi l'osservazione successiva all'ordine medio (sia quella successiva e quella dopo ancora) (come mi qui sto così).

$$q_1 = \frac{30,4 + 30,5}{2} = 30,45$$

$$q_3 = \frac{31,3 + 31,2}{2} = 31,25$$

⑧ la media $\bar{m} = \frac{0,6 + 1,2 + \dots + 0,8}{6} = 0,85$ (minuti).

Usiamo la formula "breve" $\sigma^2 = \left(\sum_{i=1}^M f_i m_i^2 \right) - \bar{m}^2$

m_i	m_i	f_i	m_i^2	$f_i m_i^2$
0,6	2	0,33	0,36	0,12
0,8	1	0,16	0,64	0,106
0,9	1	0,16	0,81	0,135
1,0	1	0,16	1	0,16
1,2	1	0,16	1,44	0,24
				<u>0,7683</u>

$$\Rightarrow \sigma^2 = 0,7683 - 0,7225 = 0,04583$$

\uparrow
m

$$\sigma = 0,21408$$

⑨ la somma dei punteggi ottenuti dai concorrenti è 489 e la media \bar{m} vale 81,5 per cui è 3, e non è quindi utilizzabile.

Calcoliamo la varianza e lo scarto quadratico medio.

Usiamo la formula lunga

$$\sigma^2 = \sum_{i=1}^M f_i (m_i - \bar{m})^2 = \frac{\sum_{i=1}^M m_i (m_i - \bar{m})^2}{N}$$

A			B			C		
m_i	$m_i - \bar{m}$	$(m_i - \bar{m})^2$	m_i	$m_i - \bar{m}$	$(m_i - \bar{m})^2$	m_i	$m_i - \bar{m}$	$(m_i - \bar{m})^2$
43	-8,5	72,25	74	-7,5	56,25	72	-9,5	90,25
46	-5,5	30,25	74	-7,5	56,25	77	-4,5	20,25
77	-4,5	20,25	78	-3,5	12,25	79	-2,5	6,25
85	3,5	12,25	84	2,5	6,25	82	0,5	0,25
88	6,5	42,25	88	6,5	42,25	84	2,5	6,25
90	8,5	72,25	91	9,5	90,25	95	13,5	182,25
		<u>249,5</u>			<u>263,5</u>			<u>305,5</u>

A	$\sigma^2 = 41,58$	$\sigma = 6,45$ ← scelgo la squadrata A.
B	$\sigma^2 = 43,92$	$\sigma = 6,63$
C	$\sigma^2 = 50,92$	$\sigma = 7,14$

14) $N = 39$ $q_0 = 21$ $q_4 = 80$

quante corse? \neq ← man è uncolante!

$$\log_2 x = \frac{\log_{10} x \cdot \ln 2}{\log_{10} 2 \cdot \ln 2}$$

Per i miei dati me scelgo 9

c_i	inf_i	sup_i	m_i	f_i	F_i	d_i
1	20	25	3	0,077	0,077	0,0154
2	25	30	8	0,205	0,282	0,041
3	30	35	12	0,307	0,589	0,0614
4	35	40	4	0,102	0,691	0,0204
5	40	45	5	0,128	0,819	0,0256
6	45	50	2	0,051	0,87	0,0102
7	50	60	1	0,025	0,895	$2,5 \cdot 10^{-3}$
8	60	70	2	0,051	0,946	$5,1 \cdot 10^{-3}$
9	70	80	2	0,051	0,997	$5,1 \cdot 10^{-3}$

↑
≈ 1 problema di

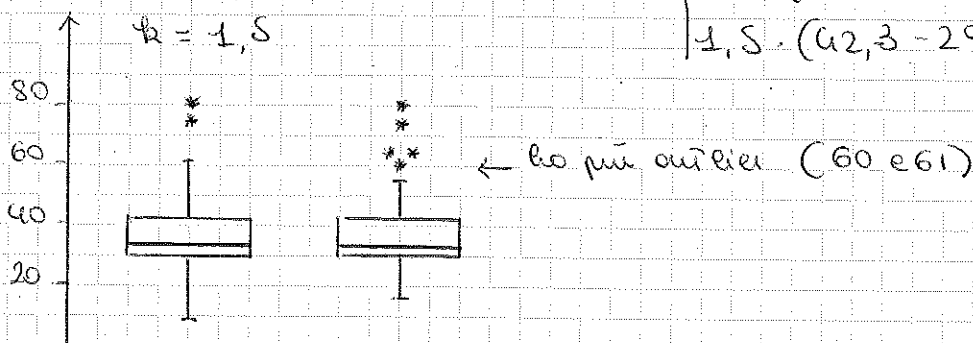
• $q_2 = inf_i + (sup_i - inf_i) \frac{0,5 - F_{i-1}}{f_i} = 30 + 5 \cdot \frac{0,5 - 0,87}{0,051} = 33,55$ autometeorio

• $q_1 = inf_i + (sup_i - inf_i) \frac{0,25 - F_{i-1}}{f_i} = 25 + 5 \cdot \frac{0,25 - 0,077}{0,205} = 29,21$

• $q_3 = inf_i + (sup_i - inf_i) \frac{0,75 - F_{i-1}}{f_i} = 40 + 5 \cdot \frac{0,75 - 0,691}{0,128} = 42,3$

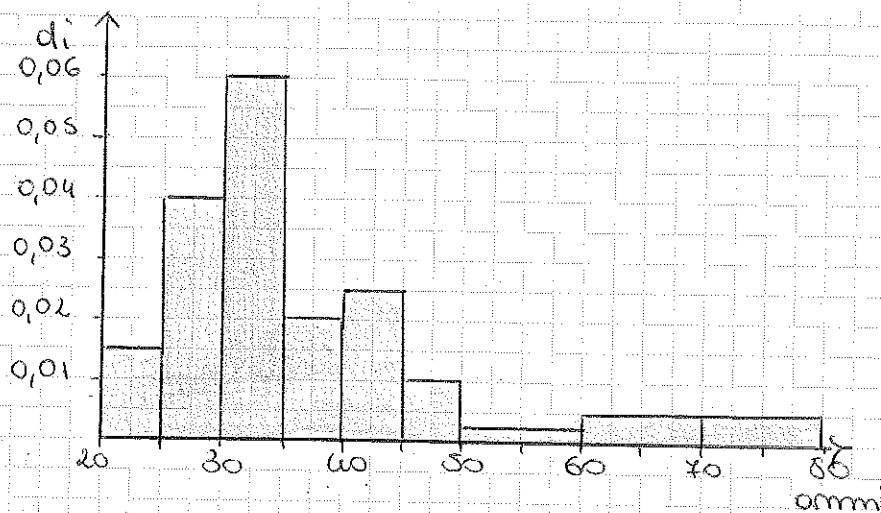
• $UA_3 = q_3 + R \cdot D = 42,3 + \begin{cases} 1 \cdot (42,3 - 29,21) = 55,1 \\ 1,5 \cdot (42,3 - 29,21) = 61,935 \end{cases}$

• $UA_1 = q_1 - R \cdot D = 29,21 - \begin{cases} 1 \cdot (42,3 - 29,21) = 16,41 \\ 1,5 \cdot (42,3 - 29,21) = 9,575 \end{cases}$



Costituisco l'istogramma dei dati → devo calcolare la densità

$$d_i = \frac{f_i}{\text{Sup}_i - \text{inf}_i}$$



faccio il confronto con la mediana calcolata sulle osservazioni e la distanza interquartile.

Ordino i dati

21 24 25 26 26 26 26 27 28 30 30 31 31 33
 33 33 34 34 34 34 35 35 37 37 38 39 41 41
 41 42 44 49 50 60 61 61 74 80
 35

calcolo la dimensione della metà $\frac{N-1}{2} = \frac{38}{2} = 19 \rightarrow$ mediana = 34
 Posto il 20° elemento

Differisce, come ci aspettavamo, dal quanto calcolato prima.

Calcolo q_1 sulle osservazioni $\frac{N-1}{4} = 9,5$ arrotondo a 10
 medio fra il 10° e l'11° elemento

$$q_1 = \frac{30+30}{2} = 30 \neq 29,21 \text{ calcolato prima}$$

Calcolo $q_3 = \frac{41+41}{2} = 41 \neq 42,3$

La distanza interquartile era $42,3 - 29,21 = 13,09$
 mentre ora è $41 - 30 = 11$

15) 1. Base: età della base
 altezza: densità ed frequenza $\frac{f_i}{\text{sup}_i - \text{inf}_i} = d_i$ $N=100$

$$f_1 = 10 \cdot 0,013 = 0,13 \rightarrow 13 p_3$$

$$f_2 = 10 \cdot 0,017 = 0,17 \rightarrow 17 p_3$$

} 30 p₃

$$f_3 = 10 \cdot 0,035 = 0,35 \rightarrow 35 p_3$$

$$f_4 = 10 \cdot 0,020 = 0,20 \rightarrow 20 p_3$$

} 55 p₃

3. Da ogni classe usiamo il valore centrale \bar{c}_i

c_i	inf_i	sup_i	\bar{c}_i	f_i	$\bar{c}_i f_i$	F_i
1	50	60	55	0,13	7,15	0,13
2	60	70	65	0,17	11,05	0,3
3	70	80	75	0,35	26,25	0,65
4	80	90	85	0,20	17	0,85
5	90	100	95	0,15	15,75	1

$$\bar{0} = \sum_{i=1}^n \bar{c}_i f_i = 77,2$$

4. moda \rightarrow scegliere classe con la frequenza maggiore.

5. Trovare classe i in cui cade q_2 e calcolo l'area residua

$$0,5 - F_{i-1} = 0,5 - 0,3 = 0,2$$

Calcolo la base del rettangolo "parziale"

$$b \frac{f_i}{\text{sup}_i - \text{inf}_i} = 0,2 \Rightarrow b = \frac{0,2 \cdot 10}{0,35} = 5,714$$

$$\Rightarrow q_2 = \text{inf}_i + b = 70 + 5,714 \approx 76 \text{ anni!}$$

6.	q_1	area residua	$0,25 - 0,13 = 0,12$	} $D = 85 - 67 = 48$
		base b	$b = \frac{0,12 \cdot 10}{0,17} = 7,06$	
		$q_1 = 60 + 7,06 \approx 67 \text{ anni!}$		
	q_3	area residua	$0,78 - 0,65 = 0,10$	}
		base b	$b = \frac{0,10 \cdot 10}{0,20} = 5$	
		$q_3 = 80 + 5 = 85 \text{ anni!}$		

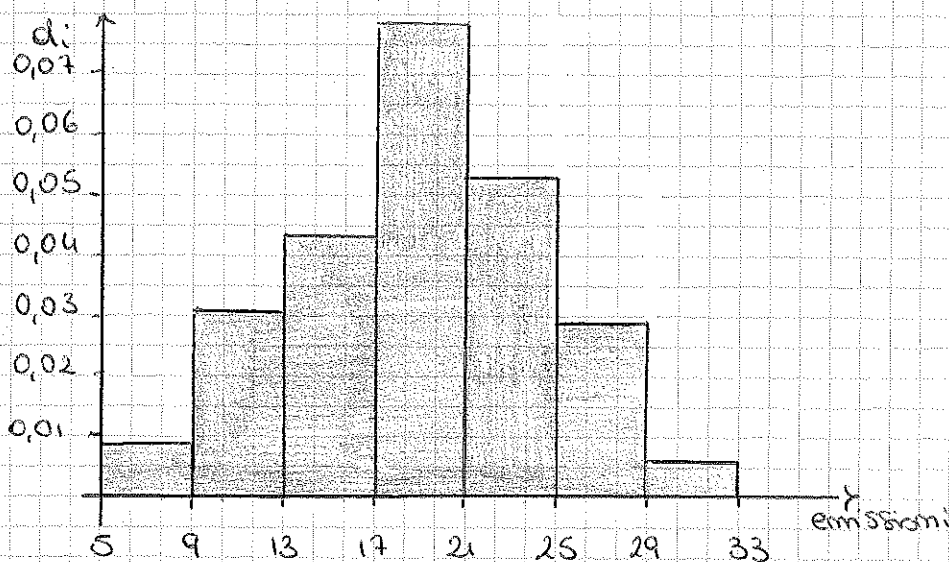
ESERCITAZIONE 1 - PROBLEMI PROPOSTI

- ②
1. quantitativo continuo.
 2. la compo di variazione o range di dati vale
 $R = 31,8 - 6,2 = 25,6$
 3. quante classi? $\sqrt{\quad}$ ← non è unicolore!

c_i	inf_i	sup_i	m_i	f_i	F_i	d_i
1	5,0	9,0	3	0,0375	0,0375	0,0094
2	9,0	13,0	10	0,1250	0,1625	0,0312
3	13,0	17,0	14	0,1750	0,3375	0,0437
4	17,0	21,0	25	0,3125	0,65	0,0781
5	21,0	25,0	17	0,2125	0,8625	0,0531
6	25,0	29,0	9	0,1125	0,975	0,0281
7	29,0	33,0	2	0,0250	1	0,0062

4.

$$d_i = \frac{f_i}{sup_i - inf_i}$$



- ③ 1. qualitativo non-accumabile.

c_i	m_i	f_i
flusso di traffico	6	0,125
intransite	22	0,438
evacuazione	43	0,271
simmetria	8	0,062
altro	5	0,104

④ 1. $\bar{m} = \frac{(3280 + 3320 + \dots + 3500 + 3780)}{20} = 3241 \text{ g}$

$M = 3255$

2. $\bar{m} = 3102 \text{ g}$

$M = 3245$

10) Per il primo strumento il coefficiente di variazione è

$$CV_1 = \frac{0,015}{3,92} \cdot 100 = 0,38\%$$

mentre per il secondo è

$$CV_2 = \frac{0,008}{1,54} \cdot 100 = 0,52\%$$

Il primo strumento è relativamente più preciso del secondo.

11)
$$\bar{m} = \frac{1}{40} (9 + 16 + 15 + 20 + 30 + 42) = 3,3$$

mediana: sarà pari alla semisomma del 20° e 21° valore (modo quindici) nella classe mediana al punteggio 3)

$$M = \frac{3+2}{2}$$

moda: punteggio al cui corrisponde la maggior frequenza

$$\hat{x} = 1$$

$q(m_i)$	m_i	f_i	m_i^2	$f_i \cdot m_i^2$
1	9	0,225	1	0,225
2	8	0,2	4	0,8
3	5	0,125	9	1,125
4	5	0,125	16	2
5	6	0,15	25	3,75
6	7	0,175	36	6,3

$$\sigma^2 = \left(\sum_{i=1}^n f_i \cdot m_i^2 \right) - \bar{m}^2$$

$$\sigma^2 = 0,225 + 0,8 + 1,125 + 2 + 3,75 + 6,3 - 10,89 = 4,31$$

$$\sigma = 1,14$$

12)
$$\bar{m} = \frac{300 + 50 + \dots + 150}{9} = 136,6$$

$$\sigma^2 = \frac{50800,04}{9} = 5644,45$$

$$\sigma = 75,13$$

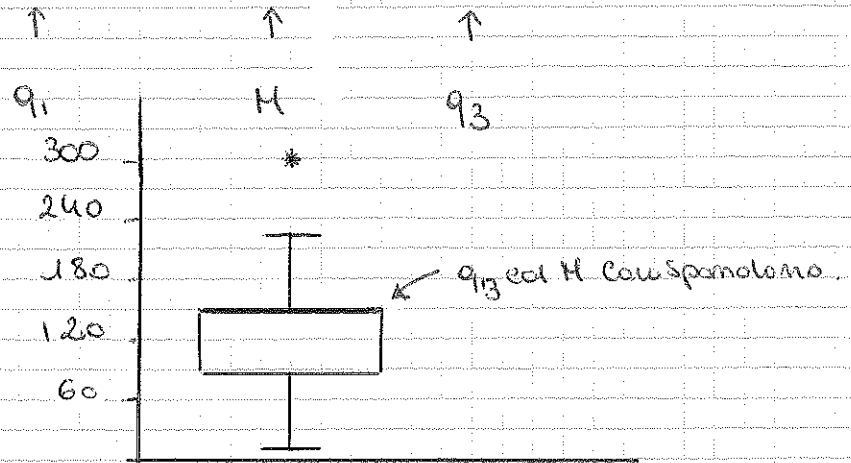
$$\sigma^2 = \frac{\sum_{i=1}^n m_i (m_i - \bar{m})^2}{N}$$

impresa)	m_i	$(m_i - \bar{m})$	$(m_i - \bar{m})^2$
1	300	163,4	699,56
2	50	-86,6	7499,56
3	200	63,4	4019,56
4	50	-86,6	7499,56
5	150	13,4	179,56
6	80	-56,6	3203,56
7	150	13,4	179,56
8	100	-36,6	1339,56
9	150	13,4	179,56

C_i	m_i
50	2
80	1
100	1
150	3
200	1
300	1

\Rightarrow mediana) $M = 150$
ordeno:

50 50 80 100 150 150 150 200 300



$\rightarrow R=1$

$$UAS = q_3 + R \cdot D =$$

$$= 150 + 1 \cdot (150 - 80) =$$

$$= 150 + 70 = 220$$

$$UAI = q_1 - R \cdot D = 80 - 70 = 10$$

Solo 300 é atípico.

(13)	C_i	uf_i	Sup_i	m_i	f_i	d_i	F_i	$f \cdot m_i^2$
	1	30	34	37	0,13	0,0325	0,13	133,12
	2	34	38	48	0,17	0,0425	0,3	220,32
	3	38	42	20	0,07	0,0175	0,37	112
	4	42	46	33	0,117	0,02925	0,487	226,512
	5	46	50	33	0,117	0,02925	0,604	269,568
	6	50	52	21	0,074	0,037	0,678	192,474
	7	52	56	26	0,092	0,023	0,77	268,272
	8	56	58	50	0,177	0,0885	0,947	575,073
	9	58	65	15	0,053	0,0075	1	200,459

$$q_2 = 46 + 4 \cdot \frac{0,5 - 0,487}{0,604} =$$

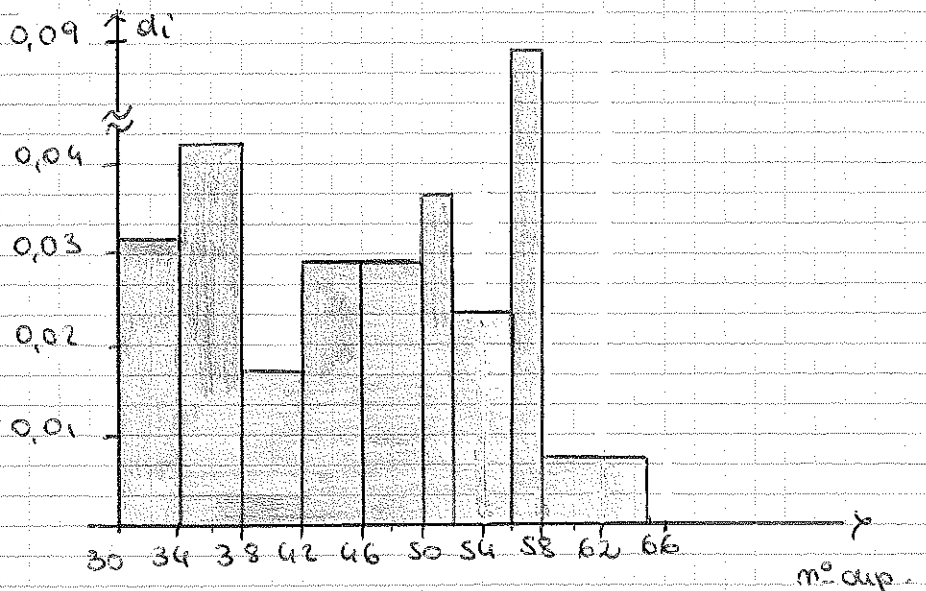
$$= 46,09$$

$$q_1 = 34 + 4 \cdot \frac{0,25 - 0,13}{0,3} =$$

$$= 35,6$$

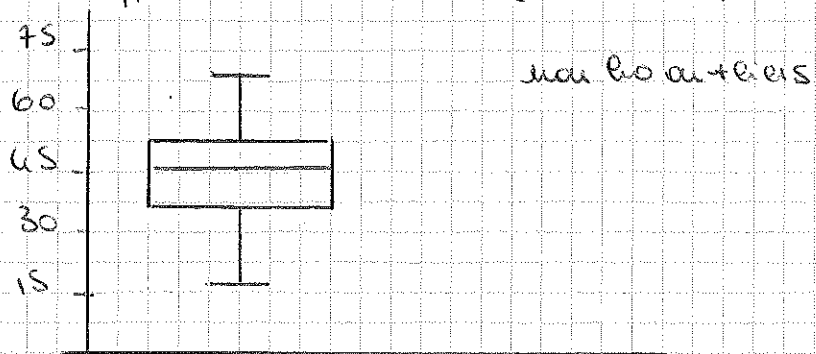
$$q_3 = 52 + 4 \cdot \frac{0,75 - 0,678}{0,77} =$$

$$= 52,37$$



$$QAS = q_3 + R \cdot D = 52,37 + (52,37 - 35,6) = 69,14$$

$$QAI = q_1 - R \cdot D = 35,6 - (52,37 - 35,6) = 18,83$$



$$\sigma^2 = 2197,798 - 2108,6464 =$$

$$= 89,1516$$

$$\sigma = 9,44$$

2. Oss é mais ou skewness do Pearson

$$\frac{\bar{m} - \text{mode}}{\sigma} = \frac{45,92 - 57}{9,44} = -1,17 \quad \text{non é simétrica}$$