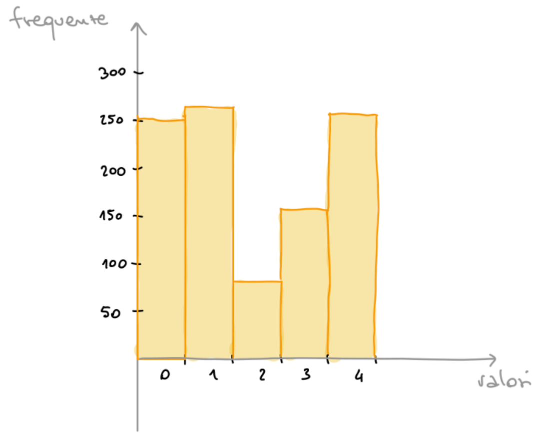
Statistica descrittiva pratica

Tuesday, 21 March 2023 15:35

1) $N = 1000$
   $\Omega = \{0, 1, 2, 3, 4\}$
   Con questi risultati:
   $f_i(0) = 251$
   $f_i(1) = 260$
   $f_i(2) = 80$
   $f_i(3) = 154$
   $f_i(4) = 255$
   a. Frequenze relative p_i
      $$p_i = \frac{f_i}{N}$$
      $$p_i(0) = \frac{251}{1000} = 0.251 = 25.1\%$$
      $p_i(1) = 0.26 = 26\%$
      $p_i(2) = 0.08 = 8\%$
      $p_o(3) = 0.154 = 15.4\%$
      $p_i(4) = 0,255 = 25,5\%$
   b. Istogramma



   c. Media, mediana, moda quartili
      $$\bar{x} = \frac{0 * 251 + 1 * 260 + 2 * 80 + 3 * 154 + 4 * 255}{1000} \approx 1.9$$
      Ora per la mediana ricordiamo del fatto che, se è pari dobbiamo fare la medi
      dispari è il valore centrale.
      $$\frac{N}{2} = \frac{1000}{2} = 500$$
      Quindi dispari
      $x_{500} + x_{501}$ $1 + 1$

$$m = \frac{x_{500} + x_{501}}{2} = \frac{1+1}{2} = 1$$

Moda $= 1$

Quartini:

$$q_1 = \frac{1}{4} \rightarrow k = 100 * p = 100 * \frac{1}{4} = 25\%$$

$$p_2 = m$$

$$p_3 = \frac{3}{4} \rightarrow k = 100 * \frac{3}{4} = 75\%$$

Ora si fa come per la mediana lol

$$q_1 = \frac{x_{250} + x_{251}}{2} = 0$$

$$q_2 = m = 1$$

$$q_3 = x_{750} = \frac{x_{750} + x_{751}}{2} = \frac{4+4}{2} = 4$$

d. Varianza e deviazione standard

$$S^2 = \frac{1}{N-1} \sum x_i - \frac{N}{N-1} * \bar{x}^2$$

$$\bar{x} = 1.902$$

Quindi possiamo calcolare

$$= \frac{1}{999}(0^2 * 251 + 1^2 * 260 + 2^2 * 80 + 3^2 * 154 + 4^2 * 255) - \frac{1000}{999}($$

$$\simeq 2.431$$

Calcoliamo deviazione standard

$$s = \sqrt{s^2} \simeq 1.56$$

2) I dati riportati misurano la densità della terra

5.5	5.61	5.88	5.07	5.26	5.55	5.36	5.29	5.58
5.65	5.57	5.53	5.62	5.29	5.44	5.34	5.79	5.1
5.27	5.39	5.42	5.47	5.63	5.34	5.46	5.3	5.75
5.68	5.85							

Possiamo notare che non abbiamo dati distinti, quindi per poterli studiare

Dobbiamo per forza lavorare per classi

$N = 29, x_{\min=5.07}, x_{max} = 5.88$

Scegliamo di raggruppare ogni 0.15, come si sceglie? Tira i dadi

$(5.00, 5.15] \rightarrow 2$

$(5.15, 5.30] \rightarrow 5$

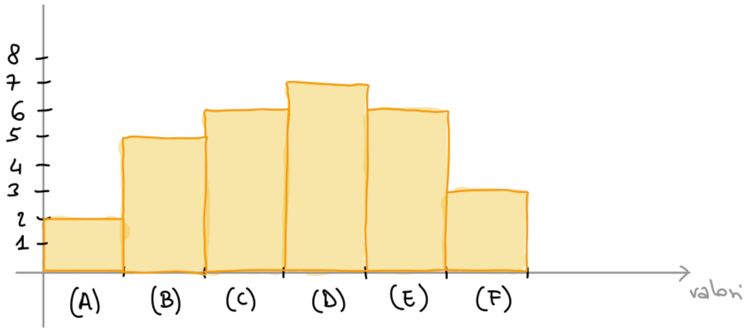
$(5.30, 5.45] \rightarrow 6$

$(5.45, 5.60] \rightarrow 7$

$(5.60, 5.75] \rightarrow 6$

$(5.75, 5.90] \rightarrow 3$

frequente



3) Dati riguardante incidenti negli anni

Anno	X=Incidenti	Y=Vittime
91	4	62
92	4	33
93	1	1
94	4	238
95	2	166

Sono sicuro che durante l'11 settembre 2002 c'è stato almeno 1 incidente con 297
Calcolare il coefficiente di correlazione tra i campioni x, y

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{(N-1)S_x S_y}$$

$\bar{x} = 3$

$\bar{y} = 100$

$$S_x^2 = \frac{1}{N-1}\sum(x_i - \bar{x})^2 = \frac{1}{5-1}[(4-3)^2 + (4-3)^2 + (1-3)^2 + (4-3)^2 + (2$$

$$= 2$$

$$S_y^2 = \frac{1}{N-1}\sum(y_i - \bar{y})^2 = 9784{,}5 = \frac{19567}{2}$$

$S_x = \sqrt{2} \simeq 1.41$

$S_y = \sqrt{9783{,}5} \simeq 98{,}91$

Quindi ora abbiamo tutti i dati per poter calcolare r

$$r = \frac{1}{(5-1)\sqrt{2} * \sqrt{\frac{19567}{2}}} *$$

(la formula è troppo gr

$$[(4-3)(62-100) + (4-3)(33-100) + (1-3)(1-10$$
$$+ (4-3)(238-100). + (2-3)(166-100)]$$

$\simeq 0.295$

E siccome $|r| \lesssim 0.3 \rightarrow$ Correlazione debole

4) Seguenti dati:

Anno	x = # Accusati	y = # Condannati
1825	6652	4037
1826	6988	4348
1827	6929	4236
1828	7396	4551
1829	7373	4475
1830	6962	4130

- Calcolare

  $$\bar{x} = \frac{6652 + 6988 + \cdots}{6} = 7050$$

  $$S_x^2 = \frac{1}{N-1} \sum (x - \bar{x})^2 = \frac{408678}{5}$$

  $$S_x = \sqrt{\frac{408678}{5}}$$

  Ora per iniziare a fare i quartili
  Dobbiamo prima ordinare
  6652, 6929, 6962, 6988, 7373, 7396
  $N = 6$
  Quindi

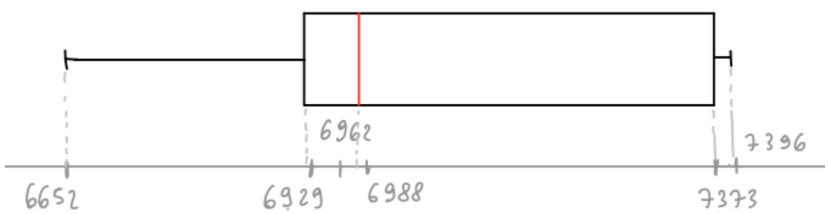
  $$m = \frac{x_3 + x_4}{2} = \frac{6962 + 6988}{2} = 6975$$

  $$q_1 = \frac{1}{4} * 6 = \frac{3}{2} = 1.5 \sim 2 \rightarrow x_2 = 6929$$

  $$q_3 = \frac{3}{4} * 6 = \frac{3}{2} * 3 = 4.5 \sim 5 \rightarrow x_5 = 7373$$

  $$\Delta = q_3 - q_1 = 444$$

  Box plot:



- Coefficiente di correlazione tra X e Y = r

  $$r = \frac{\sum x_i y_i - N * \bar{x} * \bar{y}}{(N-1) S_x S_y}$$

  Saranno tanti calcoli (che io non farò

  $\sqrt{408678}$

- $S_x = \sqrt{\frac{\text{[illegible]}}{5}}$
  $\bar{x} = 7050$
  (Calcolato prima)
- $S_y^2 = \frac{1}{N-1} \sum (y - \bar{y})^2$
  - $\bar{y} = \frac{4037 + \cdots}{6} = \frac{25777}{6}$

  $$S_y^2 = \frac{1}{N-1} * \sum (y_u - \bar{y})^2$$
  $$= \frac{7052286}{180}$$
  $$S_y = \sqrt{\frac{7052286}{180}}$$
- $\sum x_i y_i = \cdots$

  E nulla non mi metto a ricopiare ciò che ha fatto la prof
- Scrivere tabella delle frequenze assoluta e relativa del campion x
  E calcolare la media usando solo la tabella

  Allora, inanzitutto tutte le frequenze assolute F_i=1
  Visto che non abbiamo nessuna ripetizione
  La frequenza relativa invece è
  $\frac{F_i}{N} = \frac{1}{6} \forall x$
  Ora
  $$\bar{x} = \sum x_i * \frac{f_i}{p_i} = \frac{6652 * 1}{6} + \frac{6988 * 1}{6} + \cdots \sim \frac{1}{N} \sum x_i$$

5) Dati i seguenti dati

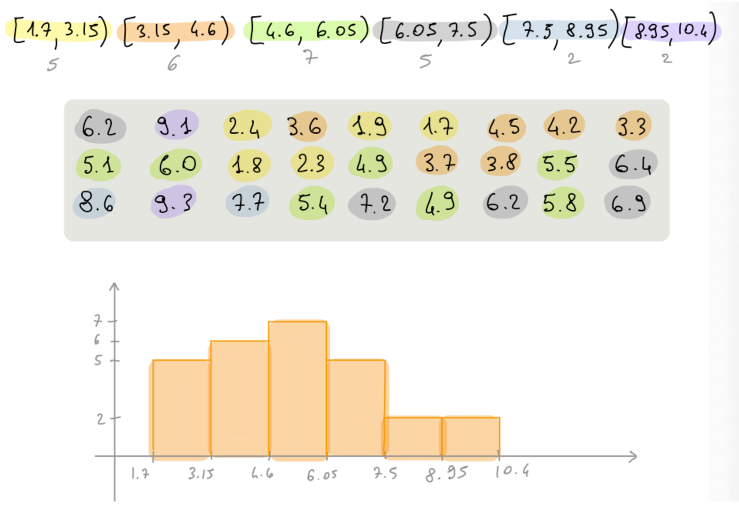
6.2	9.1	2.4	3.6	1.9	1.7	4.5	4.2	3.3
5.1	6.0	1.8	2.3	4.9	3.7	3.8	5.5	6.4
8.6	9.3	7.7	5.4	7.2	4.9	6.2	5.8	6.9

Dividerli in classe e fare istogramma frequenze
Allora (prima li divido come li dividerei io, poi mostro come l'ha fatto la prof)
$min = 1.7$
$max = 9.3$
$N = 27$
$9.3 - 1.8 \quad 7.5$

$\frac{\ldots}{10} = \frac{\ldots}{10} = 0.75$

$[1.7, 2.55], [2.55, 3.3], \ldots$

Ora metto la soluzione della prof



Perché non lo sto facendo io? SONO PIGRO

6) $1, 3, 3, 8, z$

a. Esprimere media e mediana in funzione di z

$$\bar{x} = \frac{1+3+3+8+z}{5} = \frac{15}{5} + \frac{z}{5} = 3 + \frac{z}{5}$$

$N = 5$

Mh

Mhhhhh

Z1338 = 3

1z338=3

133z8=3

1338z=3

Quindi

$m = 3\ \forall z \in R$

b. Per quali $q_3 = 6$

Da notare che, noi non abbiamo un valore = 6 quindi

$z = 6$

13368

Quando z=6 $q_3 = 6$

E solo quando z=6

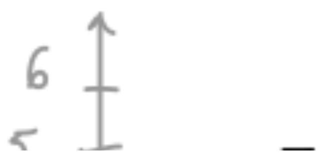
c. $q_1 = 1$

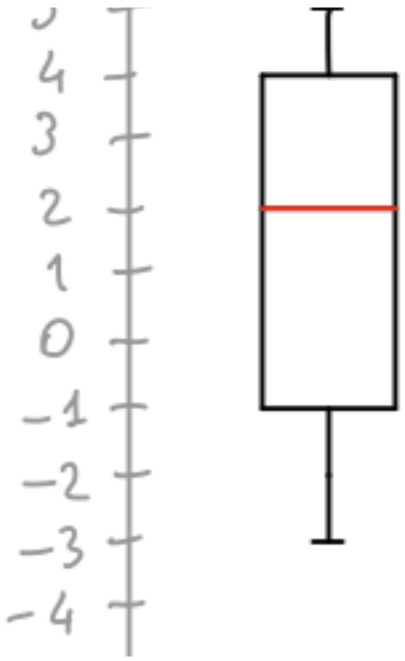
Allora

Z1338

Quindi $z \leq 1$

7) Si consideri seguente box plot





a. Quanto vale Δ

   $\Delta = q_3 - q_1 = 4 - (-1) = 5$

b. Media e mediana coincidono?

   A me non sembra proprio, vado però ad intuito

c. Posso affermare che $S_x^2 = 1$

   Questo vuol dire che la dispersione è 1

   Però cioè, guardandolo la dispersione si nota che è elevata

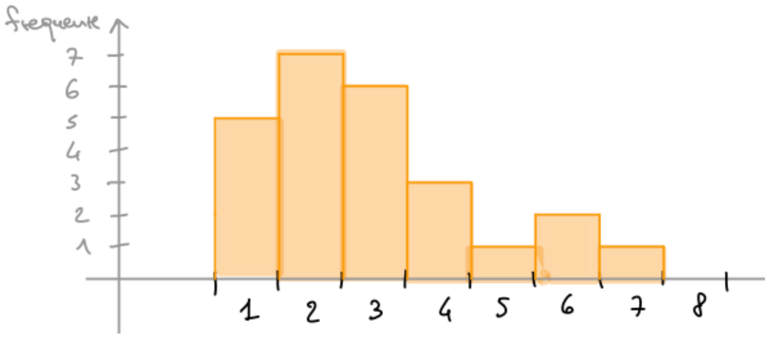
d. Range $x_1, \dots, x_n$

   Ma che razza di domanda è?

   E' così tanto facile che ti fa venire i dubbi esistenziali LOL

   $[5, -3]$

8) Dato



Calcolare media

Beh,

$N = 5 + 7 + 6 + 3 + 1 + 2 + 1 = 25$

$$\bar{x} = \frac{1 * 5 + 2 * 7 + 3 * 6 + 4 * 3 + 1 * 5 + 6 * 2 + 7 * 1}{25} = 2.92$$