# CORSO DI STATISTICA (parte 1) - ESERCITAZIONE 3

*Dott.ssa Antonella Costanzo*

*a.costanzo@unicas.it*

**Indici di posizione per caratteri quantitativi discreti e continui**

Il seguente data set riporta la rilevazione di alcuni caratteri su un collettivo di 20 studenti.

| Studente | Sesso | Età | Red | Istituto di provenienza | Voto al diploma | Statura (cm) | Colore occhi | Voto esame | Giud. sul corso |
|---|---|---|---|---|---|---|---|---|---|
| 1 | M | 22 | 0,7 | ITC | 96 | 173 | Nero | 26 | Pessimo |
| 2 | F | 20 | 0,2 | Liceo Classico | 92 | 168 | Marrone | 26 | Ottimo |
| 3 | F | 30 | 1,6 | Liceo Classico | 90 | 165 | Marrone | 30 | Buono |
| 4 | M | 22 | 2,5 | Liceo Scient | 85 | 180 | Nero | 25 | Buono |
| 5 | F | 26 | 3,2 | ITI | 100 | 163 | Azzurro | 30 | Pessimo |
| 6 | F | 20 | 0,5 | ITC | 74 | 160 | Nero | 24 | Pessimo |
| 7 | M | 26 | 4,2 | Liceo Scient | 60 | 177 | Marrone | 20 | Suff |
| 8 | M | 30 | 1,3 | ITC | 76 | 164 | Verde | 18 | Ottimo |
| 9 | F | 27 | 1,2 | Liceo Scient | 100 | 158 | Azzurro | 29 | Ottimo |
| 10 | F | 25 | 1,7 | ITI | 95 | 170 | Nero | 25 | Pessimo |
| 11 | F | 25 | 1,9 | ITI | 85 | 167 | Nero | 25 | Buono |
| 12 | M | 22 | 0,7 | ITC | 97 | 159 | Marrone | 27 | Buono |
| 13 | F | 21 | 0,4 | Liceo Classico | 65 | 174 | Azzurro | 21 | Ottimo |
| 14 | F | 24 | 1,8 | Liceo Scient | 70 | 164 | Verde | 30 | Suff |
| 15 | M | 20 | 1,9 | Liceo Scient | 80 | 177 | Nero | 28 | Suff |
| 16 | F | 21 | 3,2 | Liceo Classico | 93 | 172 | Nero | 27 | Pessimo |
| 17 | F | 27 | 2,1 | ITC | 100 | 166 | Marrone | 26 | Suff |
| 18 | F | 22 | 0,1 | ITI | 84 | 160 | Marrone | 24 | Buono |
| 19 | M | 23 | 1,6 | Liceo Scient | 92 | 170 | Azzurro | 27 | Ottimo |
| 20 | F | 23 | 2,2 | Liceo Scient | 73 | 184 | Verde | 23 | Buono |

Nel nostro dataset:

Caratteri quantitativi discreti: Voto al diploma, Voto all'esame

Caratteri quantitativi continui: Età, Reddito, Statura

Quesiti:

1. Determinare la moda, la media aritmetica, la mediana, Q1, Q3, 12-esimo percentile, 83-esimo percentile, per il carattere Voto all'esame organizzato in tabella di frequenze.
2. Determinare la moda, la media aritmetica, la mediana per il carattere Statura organizzato in classi equiampie (ampiezza 10).
3. Costruire la distribuzione di frequenza per la variabile Età organizzata in 4 classi. Calcolare moda, media, mediana e quartili. Valutare qual è la frazione (percentuale) di studenti intervistati che hanno un età al più pari a 21. Valutare inoltre:
   - quanti studenti hanno un età minore di 25 oppure maggiore di 27.
   - la percentuale di studenti con un età compresa tra 24 e 28.

Ulteriori quesiti (la cui soluzione non è stata trattata a lezione ma utili per "esercitarsi").

4. Costruire la distribuzione di frequenza per la variabile Reddito suddivisa in classi. Calcolare moda, mediana, Q1 e Q3. Si decide di attribuire un buono pasto agli studenti con reddito minore di 1,5. Valutare quanti studenti (qual è la frazione / qual è la percentuale di studenti) che beneficeranno del buono pasto.
5. Consideriamo la situazione descritta al punto 4 e supponiamo che si decida di cambiare la soglia minima di reddito che attribuisce il diritto al buono pasto: la nuova soglia di reddito viene fissata a 1 (migliaia di euro). Calcolare quanti studenti avranno diritto ad ottenere il buono pasto.

*Soluzione Q.1*

| **Voto all'esame** | $n_i$ | $f_i$ | $N_i$ | $F_i$ |
|---|---|---|---|---|
| 18 | 1 | 0.05 | 1 | 0.05 |
| 20 | 1 | 0.05 | 2 | 0.1 |
| 21 | 1 | 0.05 | 3 | 0.15 |
| 23 | 1 | 0.05 | 4 | 0.20 |
| 24 | 2 | 0.1 | 6 | 0.30 |
| 25 | 4 | 0.2 | 10 | 0.50 |
| 26 | 2 | 0.1 | 12 | 0.60 |
| 27 | 3 | 0.15 | 15 | 0.75 |
| 28 | 1 | 0.05 | 16 | 0.80 |
| 29 | 1 | 0.05 | 17 | 0.85 |
| 30 | 3 | 0.15 | 20 | 1 |
| Totale (n) | 20 | 1 | | |

**Moda:** 25

**Primo Quartile**: 24

**Secondo Quartile (mediana):**25

**Terzo Quartile**: 27

**12-esimo percentile** = $x_{\left(\frac{12*1}{100}\right)}=x_{0.12} = 21$

**83-esimo percentile** = $x_{\left(\frac{83*1}{100}\right)}=x_{0.83} = 29$

**Media aritmetica:** $\mu = \frac{\sum_{i=1}^{n} x_i * n_i}{N} = \frac{18*1+20*1+21*1+\cdots+}{20} = 25.5$

*Soluzione Q.2*

**Distribuzione in classi per la variabile Statura**

| **Statura (classi)** | $n_i$ | $f_i$ | $N_i$ | $F_i$ | $c_i = \frac{X_{max} + X_{min}}{2}$ | $a_i$ | $d_{i(ass)}$ | $d_{i(rel)}$ |
|---|---|---|---|---|---|---|---|---|
| (150, 160] | 4 | 0.2 | 4 | 0.2 | 155 | 10 | 0.4 | 0.02 |
| (160, 170] | 9 | 0.45 | 13 | 0.65 | 165 | 10 | 0.9 | 0.045 |
| (170, 180] | 6 | 0.30 | 19 | 0.95 | 175 | 10 | 0.6 | 0.03 |
| (180, 190] | 1 | 0.05 | 20 | 1 | 185 | 10 | 0.1 | 0.005 |
| Totale (n) | 20 | 1 | | | | | | |

**Classe modale**: (160,170] essendo un carattere quantitativo continuo organizzato in classi bisogna individuare la modalità a cui è associata la più alta densità di frequenza (assoluta o relativa). Nota però che in questo caso le classi sono equiampie e quindi è possibile far riferimento alle sole frequenze.

La **media aritmetica** per la variabile Statura organizzata in classi è pari a:

$\mu = \sum_{i=1}^{n} \frac{c_i * n_i}{n} = \frac{155(4)+165(9)+\cdots}{20}$ 167 cm

**Mediana:** per dati quantitativi continui organizzati in classi, la determinazione della mediana avviene nel modo seguente:

1. Individuazione della classe mediana $(x_{i-1}, x_i)$: in tale classe cadrà la modalità assunta dall'unità statistica che occupa la posizione centrale della distribuzione ordinata delle modalità. In tal caso corrisponde a (160,170]

2. Il valore che assume la mediana nella classe individuata deve essere calcolato tramite l'operazione di interpolazione sulla funzione di ripartizione empirica (se si utilizzano le frequenze relative cumulate). Tale procedimento sarà utile anche per la ricerca dei quartili. Utilizzando le frequenze cumulate relative, in generale sappiamo che il particolare quantile $q_{x_i}$ può essere determinato con la seguente:

$$q_{x_i} = x_{i-1} + \frac{x_i - x_{i-1}}{f_i}\left(F\left(q_{x_i}\right) - F_{i-1}\right)$$

dove $q_{x_i} = Me$ nel nostro caso per cui $F\left(q_{x_i}\right) = 0.5$

Avremo quindi:

$$q_{x_i} = 160 + \frac{170 - 160}{0.45}(0.5 - 0.2) = 166.66$$

Nota: $f_i = F_i - F_{i-1}$

L' espressione utilizzata per calcolare la mediana (e in generale l'i-esimo quantile) per un carattere organizzato in classi è il risultato della seguente proporzione:

$q_{x_i}$=?

$x_{i-1}$ $\quad$ *F*($q_{x_i}$)=0.5 $\quad$ $x_i$

$F_{i-1}$ $\quad$ $F_i$

$$(x_i - x_{i-1}):(q_{x_i} - x_{i-1}) = (F_i - F_{i-1}):(F(q_{x_i}) - F_i)$$

$$(q_{x_i} - x_{i-1}) = \frac{(x_i - x_{i-1})(F(q_{x_i}) - F_i)}{(F_i - F_{i-1})}$$

$$q_{x_i} = x_{i-1} + \frac{(x_i - x_{i-1})(F(q_{x_i}) - F_i)}{(F_i - F_{i-1})} \equiv q_{x_i} = x_{i-1} + \frac{x_i - x_{i-1}}{f_i}(F(q_{x_i}) - F_{i-1})$$

Nel nostro caso avremo quindi:

$$q_{x_i} = Me = 160 + (170 - 160)\frac{\mathbf{0.5} - 0.20}{0.65 - 0.20} = 166.66$$

*Soluzione Q.3*

| **Età (classi)** | $\boldsymbol{n_i}$ | $\boldsymbol{f_i}$ | $\boldsymbol{N_i}$ | $\boldsymbol{RN_i}$ | $\boldsymbol{F_i}$ | $\boldsymbol{c_i}$ | $\boldsymbol{a_i}$ | $\boldsymbol{d_{i(ass)} = {n_i}/{a_i}}$ | $\boldsymbol{d_{i(rel)} = {f_i}/{a_i}}$ |
|---|---|---|---|---|---|---|---|---|---|
| (20,23] | 9 | 0.45 | 9 | 20 | 0.45 | 21.5 | 3 | 3 | 0.15 |
| (23,25] | 3 | 0.15 | 12 | 11 | 0.6 | 24 | 2 | 1.5 | 0.075 |
| (25,27] | 6 | 0.30 | 18 | 8 | 0.90 | 26 | 2 | 3 | 0.15 |
| (27, 30] | 2 | 0.10 | 20 | 2 | 1 | 28.5 | 3 | 0.66 | 0.03 |
| | 20 | 1 | | | | | | | |

**Classe modale***:* (20,23] e (25,27] distribuzione bimodale

**Media aritmetica:** $= \frac{(21.5*9)+(24*3)+(26*6)+(28.5*2)}{20}$=24

**Primo quartile** $Q_1$ ricade nella classe (20,23]

**Secondo quartile** $Q_2 = Me$ ricade nella classe (23,25]

**Terzo quartile** $Q_3$ ricade nella classe (25,27]

Per calcolare l'esatto valore di $Q_1$, Me, e $Q_3$ si procede con l'interpolazione es. sfruttando le frequenze relative cumulate:

*Svolgimento:*

$$Q_1 = 20 + (23 - 20)\frac{\mathbf{0.25} - 0}{0.45 - 0} = 22$$

$$Q_2 = 23 + (25 - 23)\frac{\mathbf{0.50} - 0.45}{0.60 - 0.45} = 24$$

$$Q_3 = 25 + (27 - 25)\frac{\mathbf{0.75} - 0.60}{0.90 - 0.60} = 26$$

*Nota generale.* Si possono usare entrambi gli approcci descritti: applicazione della formula diretta e inversa a seconda del problema (cfr. soluzione 2) o tramite il ragionamento in proporzione sempre a seconda del problema, il risultato non cambia.

**Percentuale di studenti intervistati con un età al più pari a 21.**

In questo caso siamo interessati alla soluzione del problema inverso: conosciamo $q_{x_i} = 21$ e vogliamo determinare il rango percentile $F(q_{x_i})$ associato a $q_{x_i} = 21$.

La formula generale è come abbiamo visto:

$$q_{x_i} = x_{i-1} + \frac{x_i - x_{i-1}}{f_i}\left(F\left(q_{x_i}\right) - F_{i-1}\right)$$

da questa si ricava la formula inversa:

$$F\left(q_{x_i}\right) = F_{i-1} + \frac{f_i}{x_i - x_{i-1}}\left(q_{x_i} - x_{i-1}\right)$$

Che corrisponde al seguente ragionamento in proporzione:

$q_{x_i}$ *noto*

$x_{i-1}$ $\quad$ $F(q_{x_i})$=? $\quad$ $x_i$

$F_{i-1}$ $\quad$ $F_i$

$$(x_i - x_{i-1}) : \left(q_{x_i} - x_{i-1}\right) = (F_i - F_{i-1}) : \left(F\left(q_{x_i}\right) - F_{i-1}\right)$$

*Quindi risolvendo la proporzione otteniamo*:

$$\left(F\left(q_{x_i}\right) - F_{i-1}\right) = \frac{(F_i - F_{i-1})\left(q_{x_i} - x_{i-1}\right)}{(x_i - x_{i-1})} =$$

$$F\left(q_{x_i}\right) = F_{i-1} + \frac{(F_i - F_{i-1})\left(q_{x_i} - x_{i-1}\right)}{(x_i - x_{i-1})} \equiv F_{i-1} + \frac{f_i}{x_i - x_{i-1}}\left(q_{x_i} - x_{i-1}\right)$$

*Nel nostro caso la percentuale di studenti con età al più pari a* 21 è:

$$F\left(q_{x_i}\right) = 0 + + \frac{0.45}{23 - 20}(21 - 20) = 0.15 = 15\%$$

*Ulteriore approccio (derivato):* e' possibile ragionare in termini proporzionali direttamente in tabella sul numero di studenti che hanno un età al più pari a 21 (sfruttando le frequenza cumulate assolute associate alla classe che contiene il valore 21). E' poi possibile risalire al valore percentuale con una semplice trasformazione del risultato. In particolare:

valore noto=21 $\quad$ N(21)=?

$x_{i-1} = 20$ $\quad$ $x_i = 23$

$N_{i-1} = 0$ $\quad$ $N_i = 9$

$$(N_i - N_{i-1}) : (x_i - x_{i-1}) = N(21) : (21 - x_{i-1})$$

$$(9 - 0) : (23 - 20) = N(21) : (21 - 20)$$

$$N(21) = \frac{9 * 1}{3} = 3$$

Solamente 3 studenti hanno un età al più pari a 21. Il corrispondente rango percentile sarà:

$$\frac{N(21) * 100}{n} = \frac{3 * 100}{20} = 0.15 = 15\%$$

**Numero di studenti con un età minore di 25 oppure maggiore di 27:**

Un modo agevole di procedere è "spezzare" il problema:

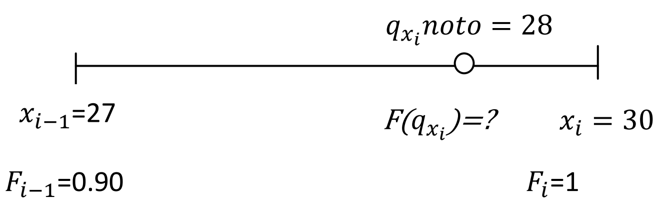
1. il numero di studenti con età minore di è pari a N(25)=12 (soluzione banale)
2. il numero di studenti con età maggiore di 27 è data dal complemento rispetto numero di studenti con età minore di 27. In particolare, il numero di studenti con età minore di 27 è pari a N(27)=18 (anche qui soluzione banale) per cui quelli con età maggiore di 27 sono n-N(27)=20-18=2. In realtà si può ragionare più facilmente a partire dalle frequenze assolute retro cumulate che indicano il numero di unità che presentano un valore del carattere maggiore o uguale ad una certa modalità RN(27)=2.

Il risultato finale è dato da: N(25) +(n-N(27))= 12+2=14
oppure, in maniera del tutto equivalente, utilizzando direttamente le frequenze retro cumulate: N(25)+RN(27)=12+2=14

**Percentuale di studenti con età compresa tra 24 e 28**

Vuol dire, in termini pratici, individuare la percentuale di studenti con età almeno pari a 24 e non più grande di 28. Una possibile (e veloce) strategia in questo caso è:

1. individuare la percentuale di studenti con età al più pari a 28.




Impostando la seguente proporzione

$(x_i - x_{i-1}) : (q_{x_i} - x_{i-1}) = (F_i - F_{i-1}) : (F(q_{x_i}) - F_{i-1})$

$(30 - 27) : (28 - 27) = (1 - 0.90) : (F(28) - 0.90)$

e risolvendo per l'incognita $F(28)$ avremo:

$F(28) = 0.90 + \frac{1*0.10}{3} = 0.933$ ossia 93.3 % degli studenti con età al più pari a 28

2. calcolare quindi la percentuale di soggetti con età al più pari a 24 e "sottrarre" questo valore alla percentuale complessiva di studenti con un età al più pari a 28 (individuata al punto 1).
1. individuare la percentuale di studenti con età al più pari a 28.

$q_{x_i} noto = 24$

$x_{i-1}$=23 $\quad$ $F(q_{x_i})$=? $\quad$ $x_i = 25$

$F_{i-1}$=0.45 $\quad$ $F_i$=0.6

Impostando la seguente proporzione

$(x_i - x_{i-1}) \colon (q_{x_i} - x_{i-1}) = (F_i - F_{i-1}) \colon (F(q_{x_i}) - F_{i-1})$

$(25 - 23) \colon (24 - 23) = (0.6 - 0.45) \colon (F(24) - 0.45)$

e risolvendo per l'incognita $F(24)$ avremo:

$F(24) = 0.45 + \frac{1*0.15}{2} = 0.525$ ossia 52.5 % degli studenti con età al più pari a 24

A questo punto "escludo" (e quindi sottraggo) dalla percentuale complessiva di studenti con al più 28 anni la percentuale di tutti quelli che sono più piccoli di 24 anni per cui 0.933-0.525=0.408.
La percentuale di studenti con un età compresa tra 24 e 28 è del 40.8%.

*Alternativamente (ragionamento speculare):*

Percentuale di studenti con età maggiore di 24= 1-%studenti con età al più pari a 24=1-0.525=0.475

Percentuale di studenti con età maggiore di 28=1- % studenti con età minore di 28=1-0.933=0.067

Dalla proporzione totale degli studenti che hanno più di 24 anni, siccome devo includere al massimo studenti che hanno 28 anni, devo "togliere" la proporzione di studenti che hanno un età maggiore di 28 che è data dal complemento a 1 della percentuale di studenti con al massimo 28 anni. Avremo quindi: 0.475-0.067=0.408

*Soluzione Q.4*

| Reddito (classi) | $n_i$ | $f$ | $N_i$ | $F_i$ | $P_i$ | $c_i$ | $a_i$ | $d_{i(ass)}$ |
|---|---|---|---|---|---|---|---|---|
| [0.1, 1) | 6 | 0.3 | 6 | 0.3 | 30 | 0.55 | 0.9 | 6.6 |
| [1,1.9) | 6 | 0.3 | 12 | 0.6 | 60 | 1.45 | 0.9 | 6.6 |
| [1.9, 4.2] | 8 | 0.4 | 20 | 1 | 100 | 3.05 | 2.3 | 3.48 |
| Totale (n) | 20 | 1 | | | | | | |

**Classe modale**=[0.1, 1) e [1, 1.9) distribuzione bimodale

**Q1**=0.85

**Me=Q2**=1.60

**Q3**=2.76

Per verificare quanti individui si trovano sotto la soglia di 1,5 (migliaia di euro) che attribuisce loro il diritto ad ottenere un buono pasto occorre definire:

1. la classe in cui rientrano tali soggetti [1, 1.9) quindi l'estremo inferiore della classe è $x_{i-1} = 1$ e l'estremo superiore è $x_i = 1.9$

2. il numero (totale) di soggetti con un reddito minore (o al più pari a 1.5) è pari a:

$$N(1.5) = N_{i-1} + \frac{n_i}{x_i - x_{i-1}}\left(q_{x_i} - x_{i-1}\right) = 6 + \frac{6}{0.9}(1.5 - 1) = 9.33 \approx 9$$

in termini percentili:

$$\frac{N(1.5)}{n} * 100 = \frac{9.33}{20} * 100 = 46.65 \approx 47° \, percentile$$

il 47% degli studenti avrà diritto al buono pasto.

*Soluzione Q.5*

Se la nuova soglia di reddito viene fissata a 1 (migliaia di euro) gli studenti che beneficeranno del sussidio sono esattamente pari a 6 ovvero il 30% del totale degli studenti (soluzione banale).