CORSO DI STATISTICA (parte 1) - ESERCITAZIONE 2

Dott.ssa Antonella Costanzo

a.costanzo@unicas.it

TIPI DI MEDIA: GEOMETRICA, QUADRATICA, ARMONICA

Esercizio 1.

Uno scommettitore puntando una somma iniziale pari a 2 euro, ha conseguito un capitale di 432 euro in 3 giocate successive. In particolare, ha vinto:

- Nella prima giocata, 3 volte la somma iniziale ovvero 2*3=6 euro
- Nella seconda giocata, 8 volte la somma precedentemente vinta, ovvero 6*8=48 euro
- Nella terza giocata, 9 volte la somma precedentemente vinta, ovvero 48*9=432 euro

Qual è stata la vincita media riportata?

Soluzione

La somma complessivamente vinta è ottenuta mediante una legge moltiplicativa, ossia: 2(3*8*9)=432 euro. Per cui, determinare di quante volte in media si è moltiplicato il capitale inizialmente puntato (2 euro) occorre individuare quel valore che sostituito ai fattori moltiplicativi 3,8 e 9 nella funzione di prodotto ne lascia invariato il risultato. Tale valore è la media geometrica dei valori osservati ovvero:

$$M_g = \sqrt[3]{3 * 8 * 9} = 6$$

Osservazione: se fosse stata calcolata la media aritmetica dei valori osservati:

$$\mu = \frac{3 + 8 + 9}{3} = 6.667$$

Sulla base di questo valore si potrebbe concludere che il giocatore ha vinto mediamente 6.667 volte la somma puntata. Tuttavia tale valore **non è corretto**: partendo da un capitale iniziale di 2 euro in tre giocate successive non si ottiene il capitale finale di 432 euro, se ad ogni giocata si vince 6.667 volte la posta. Considerata la tipologia del problema, la media idonea a rappresentare i valori osservati è sicuramente la media geometrica.

Esercizio 2.

6000 abitanti di un comune sono distribuiti in 3 differenti quartieri. La densità abitativa (abitanti/Km^2) ed il numero di abitanti in ciascun quartiere sono riportati nella seguente tabella:

densità abitativa	Nr.abitanti
20	1600
30	2100
46	2300
Totale	6000

I dati rilevati forniscono la seguente informazione: nel primo quartiere vi sono 1600 abitanti, ognuno dei quali occupa 1/20-esimo di chilometri quadrati di superficie; nel secondo quartiere ve ne sono 2100 e ciascuno occupa 1/30-esimo di chilometri quadrati e infine nel terzo quartiere ve ne sono 2300 e ciascuno occupa un 1/46-esimo di chilometri quadrati di superficie.

- Calcolare la superficie totale del comune considerato:

Sol.

$$\frac{1}{20}(1600)+\frac{1}{30}(2100)+\frac{1}{46}(2300)=200\ km^2$$

- Determinare la densità abitativa media

Sol.

La densità abitativa media è pari alla media armonica ponderata delle densità abitative osservate, ovvero:

$$M_g=\frac{6000}{\frac{1}{20}(1600)+\frac{1}{30}(2100)+\frac{1}{46}(2300)}=30$$

Si può affermare che in media i quartieri del comune considerato hanno una densità abitativa pari a 30 abitanti per km^2 di superficie

Esercizio 3.

Si calcolino la media armonica e la media quadratica della seguente distribuzione:

X	n_i
5	20
20	40
40	60
60	50
80	30
Totale N	200

Soluzione

X	n_i	x_i^{-1}	$x_i^{-1}n_i$	x_i^2	$x_i^2 n_i$	$\log(x_i)$	$\log(x_i)n_i$
5	20	0.2	4	25	500	1.6094	32.1887
20	40	0.05	2	400	16000	2.9957	119.8293
40	60	0.025	1.5	1600	96000	3.6889	221.3328
60	50	0.01667	0.8333	3600	180000	4.0943	204.7172
80	30	0.0111	0.3333	8100	243000	4.4998	134.9943
Totale N	200		8.6667		535500		713.0623

$$M_A = \frac{N}{\sum_{i=1}^{5} x_i^{-1} n_i} = \frac{200}{8.6667} = 23.0769$$

$$M_Q = \sqrt{\frac{1}{N}\sum_{i=1}^{5} x_i^2 n_i} = \sqrt{\frac{1}{200}(535500)} = 51.7446$$

Si calcolino inoltre la media aritmetica e la media geometrica verificando numericamente la relazione di ordinamento:

$$M_A \leq M_g \leq \mu \leq M_q$$

$$\mu = \frac{1}{200}\sum_{i=1}^{5} x_i n_i = \frac{1}{200}(9000) = 45$$

Ricorda: proprietà media geometrica

Il logaritmo della media geometrica è la media aritmetica dei logaritmi dei valori della variabile X oggetto di studio, ossia:

$$\log(M_g) = \frac{\log(x_1)^{n_1} + \log(x_2)^{n_2} + \cdots + \log(x_n)^{n_n}}{N} = \frac{n_1 log x_1 + n_2 log x_2 + \cdots + n_n log x_n}{N}$$

$$M_g = \exp\left[\frac{1}{200}\sum_{i=1}^{5}\log(x_i)\, n_i\right] = \exp\left(\frac{713.0623}{200}\right) = \exp(3.5653) = 35.3505$$

INDICI DI POSIZIONE PER CARATTERI QUANTITATIVI DISCRETI E CONTINUI

Esercizio 4. Verifica della proprietà associativa della media aritmetica, moda, mediana e quantili per dati discreti e continui

La tabella seguente mostra i risultati relative ad un campione di 11 persone che frequentano un centro sportivo.

Iscrizione (A=Annuale, M=Mensile, S=Semestrale) - Età - Giorni (Numero di giorni di presenza a settimana) - Sesso(M=Maschio, F=Femmina), Punteggio 400 SL (Punteggio ottenuto in una gara di nuoto 400 SL per gli 11 soggetti intervistati)

Iscrizione	Statura (cm)	Età	Giorni	Sesso	Punteggio 400 SL
M	155	18	1	F	4.8
M	163	23	2	M	5.2
A	178	23	3	M	6.01
S	186	28	4	F	7.04
M	190	30	2	F	7.5
A	178	19	4	M	8.03
A	158	23	2	F	9.01
S	165	22	2	M	12.30
S	166	18	4	M	12.90
A	165	22	2	F	13.2
M	177	25	1	M	14.1

a) Dopo aver organizzato la variabile Età in classi [18,21) [21,24) [24,27) [27,31] rappresentarla graficamente, costruire la funzione di ripartizione empirica

b) Confrontare l'età media dei maschi e delle femmine che frequentano il centro sportivo a partire dai dati originali. Dimostrare la proprietà di associatività della media aritmetica

c) Determinare la moda della variabile Iscrizione

d) Calcolare media, mediana, Q1, Q3, e il 15-esimo percentile per la variabile Punteggio ai 400 SL

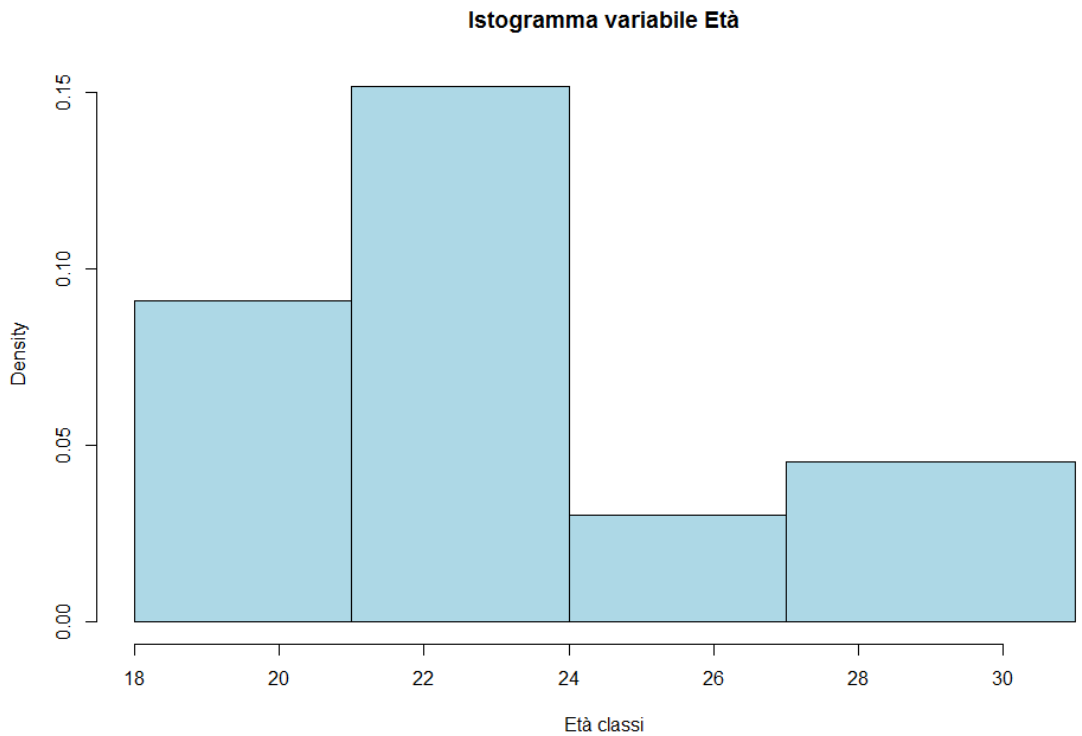
e) Calcolare moda media mediana Q1 e Q3 della variabile Età organizzata in classi.

Soluzioni

a) Indichiamo con X la variabile Età. La sua distribuzione di frequenze è pari a:

X=Età	c_i	n_i	f_i	a_i	$d_{i(ass)}$	F_i
[18,21)		3	0.27	3	1	0.27
[21,24)		5	0.45	3	1.66	0.72
[24,27)		1	0.09	3	0.33	0.81
[27,31]		2	0.18	4	0.5	1
Totale		11	1			

Rappresentazione grafica: istogramma



Funzione di ripartizione empirica

Le frequenze cumulate consentono di costruire la funzione di ripartizione empirica, $F(x)$ che rappresenta il numero di osservazioni del fenomeno minori o uguali del valore x. E' costruita ponendo sull'asse delle ascisse i valori della variabile e sull'asse delle ordinate le frequenze cumulate assolute (o relative).

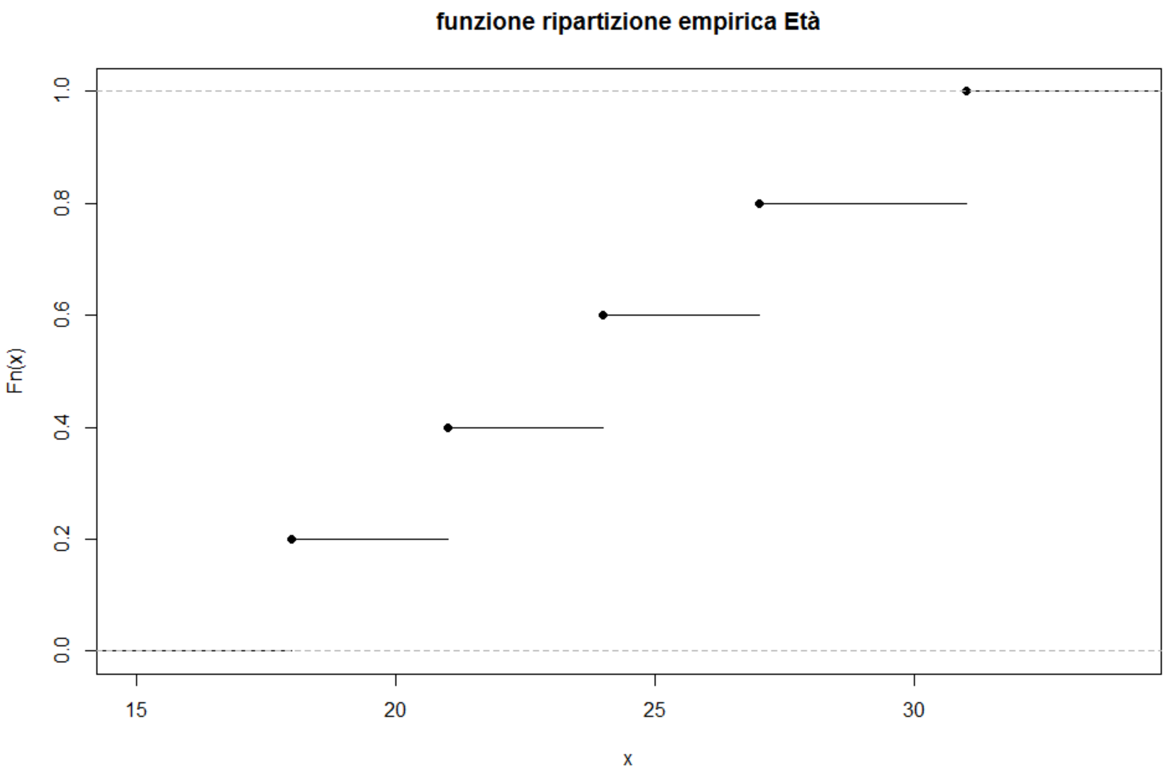
Se $\{x_1 \dots x_n\}$ sono le osservazioni (ordinate in senso crescente), con frequenze relative $f_1 \dots f_n$ la funzione di ripartizione ha espressione analitica

$$F(x) = \left\{ \begin{array}{cc} 0 & x < x_i \\ F_i = \sum_{j \le i} f_i & x_i \le x \le x_{i+1} \\ 1 & x \ge x_n \end{array} \right\}$$

In particolare per la variabile Età, la funzione di ripartizione empirica corrispondente è:

$$F(x) = \left\{ \begin{array}{r} 0 \; per \; x < 18 \\ 0.27 \; per \; 18 \le x < 21 \\ 0.72 \; per \; 21 \le x < 24 \\ 0.81 \; per \; 24 \le x < 27 \\ 1 \; per \; 27 \le x < 31 \end{array} \right\}$$

Graficamente, in caso di variabili quantitative continue organizzate in classi (es. Età) la funzione di ripartizione è rappresentata da una spezzata che unisce i punti di coordinate $(x, F(x))$.



b) Confronto tra medie, verifica proprietà associativa della media aritmetica

$$\bar{x}_F = \frac{121}{5} = 24.2 \qquad \bar{x}_M = \frac{130}{6} = 21.66$$

Verifica della proprietà associativa:

Nel caso specifico dobbiamo verificare che la media generale della variabile Età è uguale alla media ponderata delle medie di gruppo (maschi/femmine) con pesi dati dalla numerosità dei gruppi.

$\mu_{Età} = 22.81$

$$\bar{x}_F = \frac{121}{5} = 24.2 \qquad \bar{x}_M = \frac{130}{6} = 21.66$$

$$\mu_{Età} = \frac{24.2 * 5 + 21.66 * 6}{11} = 22.81$$

c) Moda per caratteri qualitativi: modalità a cui è associata la più alta frequenza (assoluta o relativa)

Iscrizione	n_i	f_i
M	4	0.37
S	3	0.26
A	4	0.37
Totale	11	1

Distribuzione bimodale: M, A

d) Mediana e quantili per caratteri quantitativi discreti (serie grezze)

Punteggio 400 SL
4.8
5.2
6.01
7.04
7.5
8.03
9.01
12.30
12.90
13.2
14.1

I quartili sono particolari percentili che dividono la distribuzione in K=4 parti uguali. In particolare:

Mediana: modalità (valore) che corrisponde alla posizione centrale della distribuzione ordinata della variabile X

Procedura per caratteri quantitativi discreti (serie grezze):

1. Ordinare in senso crescente i valori della variabile

2. Verificare la numerosità di n:

Ricordiamo che per caratteri quantitativi discreti la mediana è quel valore x tale che:

- **Se n è pari:** $\boldsymbol{Me = \frac{\left(X_{\frac{n}{2}} + X_{\frac{n}{2}+1}\right)}{2}}$

- **Se n è dispari:** $\boldsymbol{Me = X_{\frac{n+1}{2}}}$

<u>NOTA OPERATIVA:</u>indipendentemente dalla numerosità delle osservazioni (pari o dispari) è sempre possibile individuare la posizione e il valore della mediana con la regola utilizzata nel caso di n dispari e cioè $Me = X_{\frac{n+1}{2}}$

Nel nostro caso: n dispari per cui la mediana sarà

$$Me = X_{\frac{n+1}{2}} = x_{(6)} = 8.03$$

Il **primo quartile** è quel valore che lascia a sinistra il 25% dei dati e a destra il restante 75%. Per caratteri quantitativi discreti il primo quartile è quel valore x tale che:

$$\boldsymbol{Q_1} = \frac{\boldsymbol{X_{\frac{n}{4}} + X_{\frac{n}{4}+1}}}{\boldsymbol{2}} = \frac{x_{(2.75)} + x_{(3.75)}}{2} = \frac{5.2 + 6.01}{2} = 5.605$$

Il **terzo quartile** è quel valore che lascia a sinistra il 75% dei dati e a destra il restante 25%. Per caratteri quantitativi discreti il terzo quartile è quel valore x tale che:

$$\boldsymbol{Q_3} = \frac{\boldsymbol{X_{\frac{3n}{4}} + X_{\frac{3n}{4}+1}}}{\boldsymbol{2}} = \frac{x_{(8.25)} + x_{(9.25)}}{2} = \frac{12.30 + 12.90}{2} = 12.6$$

15-esimo percentile

Per il generico quantile q_i che divide la distribuzione in k parti uguali

$$q_i = x_{(\frac{i\,n}{K})}$$

- Individuo la posizione del quantile desiderato nella distribuzione ordinata di X
- A partire dalla posizione individuata distinguo la parte intera e la parte decimale; quest'ultima esprime la posizione esatta del quantile desiderato risultante dall'operazione precedente.
- Il valore del quantile di una distribuzione può essere definito come la somma pesata dei valori corrispondenti alle posizioni del quantile desiderato. I pesi sono funzione della distanza del valore del quantile rispetto alla posizione individuata.

Esempio: 15-esimo percentile, K=100 n=11 i=15

$$q_i = x_{(\frac{i\,n}{K})} = x_{(\frac{15*11}{100})} = x_{(1.65)}$$

Posizione di x= 1.65 (1 parte intera, 65 parte decimale). Notare che il valore di x si trova sostanzialmente più vicino alla seconda che non alla prima posizione della distribuzione ordinata di X. Alla prima posizione corrisponde il valore $x_{(1)} = 4.8$ mentre alla seconda posizione $x_{(2)} = 5.2$

Il valore del 15-esimo percentile sarà dato dalla somma pesata dei valori $x_{(1)} = 4.8$ e $x_{(2)} = 5.2$ con pesi in funzione della distanza del valore del quantile rispetto alle posizioni individuate:

$$(1 - 0.65)x_{(1)} + 0.65x_{(2)} = 0.35(4.8) + 0.65(5.2) = 5.06$$

Nota operativa: questa procedura è utile per il calcolo di tutti i quantili (quartili, decili e percentili)

e) Moda media e mediana per caratteri quantitativi continui organizzati in classi

X=Età	c_i	n_i	f_i	a_i	$d_{i(ass)}$	F_i
[18,21)	19.5	3	0.27	3	1	0.27
[21,24)	22.5	5	0.45	3	1.66	0.72
[24,27)	25.5	1	0.09	3	0.33	0.81
[27,31]	29	2	0.18	4	0.5	1
Totale		11	1			

Classe modale: [21,24] 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).

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

$$\mu = \sum_{i=1}^{n} \frac{c_i * n_i}{n} = \frac{19.5(3)+22.5(5)+\cdots}{11} = 23.13$$

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 (21,24]

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:

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

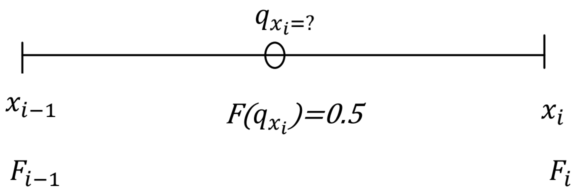
dove $q_{x_i} = Me$ nel nostro caso per cui $F\left(q_{x_i}\right) = 0.5$

Avremo quindi:

$$q_{x_i} = 21 + \frac{24-21}{0.72}(0.5 - 0.27) = 21.95 \approx 22$$

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:



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

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

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

Nel nostro caso avremo quindi:

$$q_{x_i} = Me = 21 + (24 - 21)\frac{\mathbf{0.5} - 0.27}{0.72 - 0.27} = 22.95$$