Teoria: Pag.338, 339, 342, 343, 346 esercizi pag. 362, 363, 367

1. I DATI STATISTICI

POPOLAZIONE, CARATTERE, FREQUENZA

- **Popolazione**: insieme di persone o oggetti sui quali si effettua un'indagine statistica.
- **Carattere**: caratteristica distintiva di ciascun elemento (**unità statistica**) di una popolazione statistica. È descritto mediante le **modalità** con cui esso si può manifestare:
  - **carattere quantitativo**: carattere le cui modalità sono espresse numericamente, può essere:
    - **continuo**: se può assumere gli infiniti valori di un intervallo reale; per esempio, il carattere $x \in \mathbb{R}$, dove $\mathbb{R}$ è l'insieme dei numeri reali.
    - **discreto**: se può assumere un numero finito di valori o al più una infinità numerabile; per esempio, il carattere $n \in A$, dove $A = \{1, 3, 7\}$.
  - **carattere qualitativo**: carattere le cui modalità sono espresse con parole; per esempio, il carattere «sesso» ha due modalità: «maschile» e «femminile».
- **Frequenza assoluta**: è il numero di volte in cui si presenta una modalità in una distribuzione di dati.
- **Frequenza relativa**: è il rapporto tra la frequenza assoluta e il numero totale delle unità statistiche.
- **Frequenza cumulata**: è la somma della frequenza assoluta corrispondente a una data modalità con tutte le frequenze assolute precedenti. (Le modalità devono essere ordinate in modo crescente.)

Le serie e le seriazioni

I dati statistici possono essere rappresentati mediante tabelle.
Le tabelle come la tabella 1 che riportano nella prima colonna le modalità di un carattere *qualitativo* vengono dette **serie statistiche**. Nella seconda colonna compare o il numero delle volte con il quale si presenta (*frequenza*) o il valore (*intensità*) di un carattere quantitativo associato (per esempio il prezzo).

L'insieme delle modalità di un carattere qualitativo, alle quali associamo le loro frequenze, definisce una **mutabile statistica**.
Le tabelle che mostrano la successione dei valori che un fenomeno assume in tempi successivi sono **serie storiche**.
Le tabelle come la tabella 2 che riportano nella prima colonna un carattere *quantitativo* vengono dette **seriazioni statistiche**. Nella seconda colonna compare la frequenza, cioè il numero di volte con il quale si presenta la relativa modalità.
L'insieme delle modalità di un carattere quantitativo, alle quali associamo le loro frequenze, definisce una **variabile statistica**.

Nella tabella 2 le modalità del carattere quantitativo della prima colonna sono raggruppate in **classi**, e vengono riportate le frequenze di ogni classe.

Serie statistica

Elettrodomestici	Frequenza
apparecchi TV	7
lavatrici	10
forni a microonde	8
aspirapolvere	15
Totale	40

▲ Tabella 1

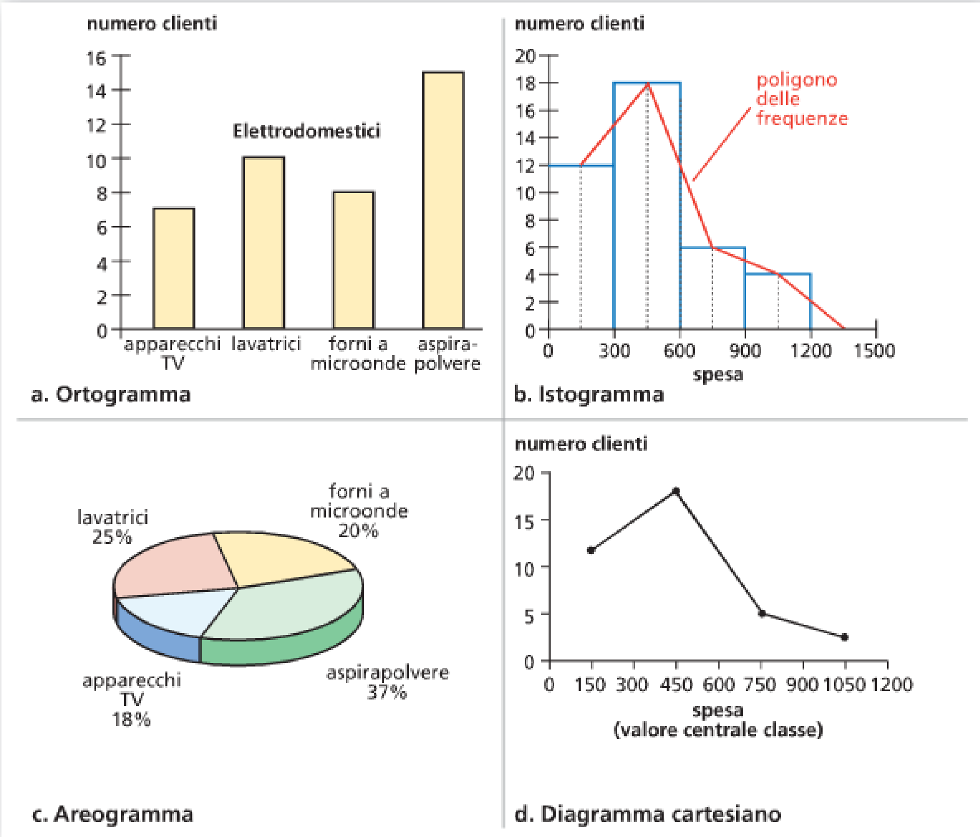
Seriazione statistica

Spesa sostenuta dai clienti (euro)	Frequenza	Frequenza relativa percentuale	Frequenza cumulata	Frequenza relativa percentuale cumulata
0-300	12	30%	12	30%
300-600	18	45%	30	75%
600-900	6	15%	36	90%
900-1200	4	10%	40	100%
Totale	40	100%		

► Tabella 2

La rappresentazione grafica dei dati

Ci sono diversi tipi di grafici per rappresentare i dati statistici e le loro frequenze. In figura 1 facciamo quattro esempi riferendoci alle tabelle precedenti.



◄ Figura 1

Un istogramma è costituito da rettangoli che hanno le basi proporzionali alle ampiezze delle classi e le aree proporzionali alle frequenze.
Se le classi *non* hanno la stessa ampiezza, le altezze dei rettangoli devono essere calcolate in modo che le aree siano proporzionali alle frequenze.
Se in un istogramma si congiungono i punti medi dei lati superiori dei rettangoli, si ottiene una spezzata chiamata **poligono delle frequenze**.

Le distribuzioni doppie di frequenze

Nella tabella 3 riportiamo i voti in italiano e in matematica di alcuni studenti.
Questa **tabella a doppia entrata** permette di conoscere quanti sono gli alunni che hanno un determinato voto in italiano e in matematica, ma anche di leggere immediatamente quanti sono gli alunni che hanno un certo voto in matematica e contemporaneamente un altro voto in italiano.

Voto in matematica \ Voto in italiano	6	7	8	9	Totale
6	4	3	0	0	7
7	1	1	1	0	3
8	1	0	1	1	3
Totale	6	4	2	1	13

▲ Tabella 3

Se vogliamo sapere quanti studenti hanno 7 in matematica e 6 in italiano, troviamo 1 all'incrocio fra la seconda riga e la prima colonna.

Si tratta di una **distribuzione doppia di frequenze** o **distribuzione congiunta** che permette l'osservazione delle unità statistiche sotto due modalità.
Quando entrambe le modalità sono quantitative si hanno **tabelle di correlazione**, se sono entrambe qualitative **tabelle di contingenza** e se una modalità è qualitativa e l'altra quantitativa **tabelle miste**.

2. GLI INDICI DI POSIZIONE CENTRALE

La media aritmetica, la mediana, la moda

La media aritmetica

DEFINIZIONE

Media aritmetica

La media aritmetica M di n numeri $x_1, x_2, \ldots, x_n$ è il quoziente fra la loro somma e il numero n.

somma dei valori

$$M = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

media aritmetica

numero dei valori

ESEMPIO

In tre test due amiche hanno ottenuto i seguenti punteggi.

Chiara:	5,4	7,8	6,3
Luisa:	4,2	8,3	6,4

Calcoliamo le medie M_C e M_L dei punteggi:

$$M_C = \frac{5{,}4 + 7{,}8 + 6{,}3}{3} = 6{,}5; \; M_L = \frac{4{,}2 + 8{,}3 + 6{,}4}{3} = 6{,}3.$$

Possiamo dire che, in media, Chiara ha ottenuto risultati migliori.

La media ponderata

DEFINIZIONE

Media aritmetica ponderata

Dati i numeri $x_1, x_2, \ldots, x_n$ e associati a essi i numeri $p_1, p_2, \ldots, p_n$, detti *pesi*, chiamiamo media aritmetica ponderata P il rapporto fra la somma dei prodotti dei numeri per i loro pesi e la somma dei pesi stessi.

somma dei prodotti dei valori per i loro pesi

$$P = \frac{x_1p_1 + x_2p_2 + \ldots + x_np_n}{p_1 + p_2 + \ldots + p_n}$$

media aritmetica ponderata

somma dei pesi

La media aritmetica può essere considerata un caso particolare di media ponderata in cui tutti i pesi sono uguali a 1.

Se calcoliamo la media aritmetica ponderata nel caso di classi, possiamo assumere come valori $x_1, x_2, \ldots, x_n$ i valori centrali di ogni classe e come pesi le frequenze. Il valore ottenuto può essere diverso dalla media aritmetica.

ESEMPIO

Calcoliamo la media aritmetica ponderata relativa alla tabella 2 (pagina 338):

$$\frac{150 \cdot 12 + 450 \cdot 18 + 750 \cdot 6 + 1050 \cdot 4}{12 + 18 + 6 + 4} = 465.$$

La media ponderata è particolarmente significativa quando i pesi servono per indicare l'*importanza* dei diversi valori.

ESEMPIO

In un quadrimestre vengono svolte prove alle quali viene attribuita una diversa importanza (compiti in classe, relazioni, interrogazioni, test). Per un certo studente i voti riportati e i pesi da attribuire ai voti sono quelli della tabella 10.
Calcoliamo la media ponderata:

$$P = \frac{5 \cdot 1 + 6 \cdot 2{,}5 + 5 \cdot 1 + 5 \cdot 1 + 7 \cdot 2{,}5 + 6 \cdot 3}{1 + 2{,}5 + 1 + 1 + 2{,}5 + 3} \simeq 5{,}95.$$

Il valore che otteniamo è maggiore di quello della media aritmetica semplice (circa 5,67), perché i voti positivi sono stati ottenuti nelle prove alle quali è stata data maggiore importanza.

Voti pesati	
Voto	**Peso**
5	1
6	2,5
5	1
5	1
7	2,5
6	3

▲ Tabella 10

La mediana

DEFINIZIONE

Mediana

Data la sequenza ordinata di n numeri $x_1, x_2, \ldots, x_n$, la mediana è:

- il valore centrale, se n è dispari;
- la media aritmetica dei due valori centrali, se n è pari.

mediana

21, 22, 26, 28, 35

21, 22, 24 26, 28

$\frac{22+26}{2} = 24$ mediana

La moda

DEFINIZIONE

Moda

Dati i numeri $x_1, x_2, \ldots, x_n$, si chiama moda il valore a cui corrisponde la frequenza massima.

$50, \underbrace{100, 100, 100}_{\text{moda}}, 200, 300, 300$

La moda indica il valore più «presente» nella distribuzione. Ci sono serie di dati che hanno più di una moda. Consideriamo i risultati di un compito in classe (tabella 11).

Voti di un compito					
Voto	4	5	6	7	8
Frequenza	2	9	3	9	2

◄ Tabella 11

La distribuzione risulta *bimodale*, avendo per moda sia 5 sia 7. Ciò significa che nella classe si possono distinguere due gruppi di studenti: uno ha ben compreso gli argomenti del compito, l'altro ha bisogno di studiarli ancora!

3. GLI INDICI DI VARIABILITÀ

Il campo di variazione, lo scarto semplice medio, la deviazione standard

Data una distribuzione di valori $x_1, \ldots, x_n$, gli **indici di variabilità** misurano la dispersione dei dati rispetto al valore medio M. I principali sono:

- **campo di variazione**: la differenza tra il valore massimo e quello minimo;

- **deviazione standard σ**: $\sigma = \sqrt{\dfrac{(x_1 - M)^2 + (x_2 - M)^2 + \ldots + (x_n - M)^2}{n}}$.

$\lvert x_i - M \rvert$	
a	*b*
$\lvert 2-7 \rvert = 5$	$\lvert 2-7 \rvert = 5$
$\lvert 3-7 \rvert = 4$	$\lvert 6-7 \rvert = 1$
$\lvert 4-7 \rvert = 3$	$\lvert 6-7 \rvert = 1$
$\lvert 4-7 \rvert = 3$	$\lvert 6-7 \rvert = 1$
$\lvert 8-7 \rvert = 1$	$\lvert 6-7 \rvert = 1$
$\lvert 8-7 \rvert = 1$	$\lvert 7-7 \rvert = 0$
$\lvert 9-7 \rvert = 2$	$\lvert 7-7 \rvert = 0$
$\lvert 9-7 \rvert = 2$	$\lvert 8-7 \rvert = 1$
$\lvert 9-7 \rvert = 2$	$\lvert 8-7 \rvert = 1$
$\lvert 14-7 \rvert = 7$	$\lvert 14-7 \rvert = 7$

▲ **Tabella 15** Scarti assoluti dalla media per le sequenze *a* e *b*.

ESEMPIO

Consideriamo le seguenti sequenze di numeri di media $M = 7$:

a. 2, 3, 4, 4, 8, 8, 9, 9, 9, 14;

b. 2, 6, 6, 6, 6, 7, 7, 8, 8, 14.

Entrambe hanno lo stesso campo di variazione 12, ma differiscono per la deviazione standard:

$$\sigma_a = \sqrt{\frac{5^2 + 4^2 + 3^2 + 3^2 + 1^2 + 1^2 + 2^2 + 2^2 + 2^2 + 7^2}{10}} = 3{,}49;$$

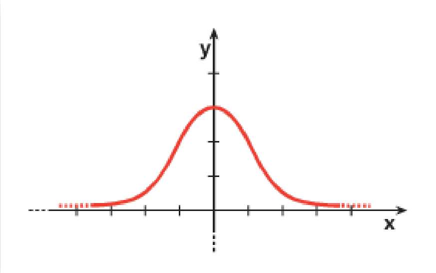
$$\sigma_b = \sqrt{\frac{5^2 + 1^2 + 1^2 + 1^2 + 1^2 + 0 + 0 + 1^2 + 1^2 + 7^2}{10}} = 2{,}83.$$

Si chiamano **distribuzioni gaussiane** (o **normali**) quelle distribuzioni il cui poligono delle frequenze ha la forma della curva di Gauss.

La distribuzione gaussiana

Il calcolo della deviazione standard assume particolare importanza nelle distribuzioni gaussiane, poiché tale indice è collegato al modo con cui le frequenze si distribuiscono intorno al valore medio.

Si può infatti dimostrare che per una distribuzione gaussiana il 68,27% dei valori è compreso fra $M - \sigma$ e $M + \sigma$, il 95,45% fra $M - 2\sigma$ e $M + 2\sigma$, e infine il 99,74% fra $M - 3\sigma$ e $M + 3\sigma$.



► **Figura 5** La curva di Gauss.

4. I RAPPORTI STATISTICI

I **rapporti statistici** sono i quozienti fra i valori di due dati statistici o di un dato statistico e di uno non statistico.
Esaminiamo i seguenti rapporti statistici: rapporti di derivazione, rapporti di densità, rapporti di composizione, rapporti di coesistenza, numeri indice.

Rapporti di derivazione: servono per confrontare due dati statistici di cui il primo deriva dal secondo.

ESEMPIO

Per comodità di lettura e interpretazione, i quozienti vengono moltiplicati per 1000.

Dati rilevati in un Comune

Anno	Numero nati	Numero morti	Popolazione
2005	980	714	76 414
2006	794	735	76 735
2007	776	705	78 018
2008	746	678	78 158
2009	731	653	78 707

Quoziente di natalità (per 1000 abitanti)	Quoziente di mortalità (per 1000 abitanti)
12,82	9,34
10,35	9,58
9,95	9,04
9,54	8,67
9,29	8,30

► Tabella 21

$$\textbf{Quoziente di natalità} = \frac{\text{numero dei nati}}{\text{popolazione}};$$

$$\textbf{Quoziente di mortalità} = \frac{\text{numero dei morti}}{\text{popolazione}}.$$

Rapporti di densità: sono i rapporti tra dati statistici e dati relativi al campo di riferimento.
Un rapporto di densità è, per esempio, il rapporto tra la popolazione e la superficie del territorio in cui abita. Un altro esempio è il rapporto tra il fatturato di un'azienda (espresso in euro) e il numero di addetti.

Rapporti di composizione: sono rapporti tra dati omogenei e servono per valutare l'importanza delle diverse modalità nella composizione del valore complessivo del fenomeno. Spesso coincidono con le frequenze relative.

ESEMPIO

L'esame della tabella 22 ci dice, per esempio, che la spesa per le manifestazioni sportive incide molto più di quella per il teatro e i concerti.

Spesa per il tempo libero

Tipo di spettacolo	Spesa
teatro e concerti	212 118
cinema	303 787
intrattenimenti vari	914 230
manifestazioni sportive	390 652
Totale	1 820 787

Rapporto di composizione in %
11,6
16,7
50,2
21,5

► Tabella 22

$$11,6\% \simeq \frac{212118}{1820787} \cdot 100, \quad 16,7\% \simeq \frac{303787}{1820787} \cdot 100, \ldots$$

Rapporti di coesistenza: sono rapporti tra le frequenze di due fenomeni diversi riferiti alle stesse unità statistiche e danno un'indicazione dello squilibrio fra dati coesistenti in uno stesso luogo o in uno stesso periodo di tempo.

Numeri indice: sono il rapporto fra un dato statistico e il valore di un dato statistico preso come elemento di riferimento (base), moltiplicato per 100. Si distinguono in numeri indice a **base fissa** e numeri indice a **base mobile**. In quest'ultimo caso, il rapporto è tra il dato statistico e quello che lo precede.

Per esempio, per confrontare i risultati di diverse scuole, si può considerare il rapporto fra il numero degli alunni respinti e quello dei promossi.

I numeri indice sono molto usati nella valutazione di serie storiche di dati.

Numeri indice a base fissa

Si divide ogni valore per quello della base fissata e si moltiplica il quoziente per 100.

ESEMPIO

Nella tabella 23 fissiamo come base il 2004 e, per indicare questa scelta, adottiamo la seguente notazione: 2004 = 100.

◀ Tabella 23

Produzione di uva		
Anno	**Produzione di uva (tonnellate)**	**Numero indice a base fissa**
2004	32,2	100,00
2005	36,8	114,29
2006	29,4	91,30
2007	32,9	102,17
2008	32,3	100,31
2009	30,2	93,79
2010	35,8	111,18

anno 2005 $\dfrac{36,8}{32,2} \cdot 100 \simeq 114,29;$

anno 2006 $\dfrac{29,4}{32,2} \cdot 100 \simeq 91,30; \ldots$

Numeri indice a base mobile

Come base si sceglie il valore che nella tabella precede il valore in esame.

ESEMPIO

◀ Tabella 24

Produzione di uva		
Anno	**Produzione di uva (tonnellate)**	**Numero indice a base mobile**
2004	32,2	n.d.
2005	36,8	114,29
2006	29,4	79,89
2007	32,9	111,90
2008	32,3	98,18
2009	30,2	93,50
2010	35,8	118,54

Il numero indice del primo anno non è determinabile in quanto non conosciamo il valore dell'anno precedente.

anno 2005 $\dfrac{36,8}{32,2} \cdot 100 \simeq 114,29;$ anno 2006 $\dfrac{29,4}{36,8} \cdot 100 \simeq 79,89; \ldots$

Notiamo che è possibile passare dai numeri indice a base fissa a quelli a base mobile dividendo un numero indice a base fissa per il suo precedente e moltiplicando il quoziente per 100.

Esercizi pag.362-363

La media aritmetica, la mediana, la moda

8 COMPLETA la seguente tabella.

Dati	Media	Mediana	Moda
3, 7, 8, 10, 3, 6, 3, 2			
12, 15, 11, 15, 19, 18, 15			

9 Nel corso del mese di giugno in un supermercato i duroni per cinque giorni sono stati offerti al prezzo di 10 euro al kg, per nove giorni al prezzo di 7 euro al kg, per quattro giorni al prezzo di 8 euro al kg e per sette giorni al prezzo di 9 euro al kg. Calcola la media aritmetica, la mediana e la moda. [8,32; 8; 7]

10 Le autovetture di un salone per la vendita di auto usate sono classificate secondo l'età dell'usato.

Età usato (mesi)	Numero autovetture
6	12
12	16
18	15
24	9
30	5
36	1
48	1
60	1

Determina la media aritmetica, la mediana e la moda. [17,4; 18; 12]

11 La seguente tabella riporta la quantità di libri venduti in una determinata settimana in 20 librerie.

Numero libri venduti	Numero librerie
50	4
60	6
70	5
80	3
90	2

Determina la media aritmetica, la moda e la mediana del numero di libri venduti. [66,5; 60; 65]

La media ponderata

13 ESERCIZIO GUIDA

Un grossista di frutta acquista quattro quantitativi di mele Golden presso aziende agricole diverse che praticano prezzi differenti. La seguente tabella espone i prezzi e le relative quantità e si vuole determinare il prezzo medio al kg.

	Azienda A	Azienda B	Azienda C	Azienda D
Prezzo (€)	0,60	0,55	0,68	0,57
Quantità (kg)	200	300	220	280

Dobbiamo calcolare una media aritmetica ponderata dei prezzi dove i pesi sono le quantità.

$$M = \frac{0,60 \cdot 200 + 0,55 \cdot 300 + 0,68 \cdot 220 + 0,57 \cdot 280}{200 + 300 + 220 + 280} \simeq 0,59.$$

14 TEST Un esame consiste in una prova di laboratorio, una prova orale e una prova scritta. Le tre prove hanno rispettivamente peso 2, 3, 5. Un candidato riceve 8 nella prova di laboratorio, 6 nella prova orale e 7 nella prova scritta. Quanto vale la media aritmetica ponderata dei punteggi?

A 6,9. B 7,2. C 6,7. D 6,5. E 7,4.

15 TEST È data la seguente tabella, relativa ai punti totalizzati giocando al tiro con l'arco.

Punti	Numero tiri
10	6
20	3
30	1
40	2

Il punteggio medio per ogni tiro è:

A 9,6.
B 25,0.
C 3,0.
D 41,5.
E 19,2.

16 La seguente tabella riporta il numero di DVD posseduti dai ragazzi di una classe.

Numero DVD	10	15	20	25
Numero ragazzi	8	7	6	3

Calcola il numero medio di DVD posseduti da ciascun ragazzo. [15,8]

Esercizi pag. 367 sul calcolo della deviazione standard
(detta anche s.q.m. cioè: scarto quadratico medio)

37 Calcola la deviazione standard delle seguenti sequenze di numeri:

a) 7; 9; 11; 13;
b) 0,9; 3,6; 9,6; 13,5; 18,9;
c) 2; 6; 10; 14; 18;
d) 3; 3; 3; 3; 3;
e) − 9; − 2; 1; 2; 3;
f) − 7; − 5; 1; 3; 8.

[a) 2,24; b) 6,53; c) 5,66; d) 0; e) 4,34; f) 5,44]

Calcola la deviazione standard delle distribuzioni descritte dalle seguenti tabelle.

38 Voti riportati da un alunno nel primo quadrimestre in inglese.

Voto	4	5	8	9
Frequenza	1	2	1	3

[2,07]

39 Temperature rilevate nel corso di una giornata invernale (espresse in °C).

Temperatura	−3	−2	2	3
Frequenza	2	3	3	2

[2,45]

40 Dalla produzione e vendita di articoli di pelletteria, una ditta, in sei mesi successivi, ha ottenuto i seguenti guadagni in euro: 100 000, 125 000, 140 000, 135 000, 160 000, 110 000. Calcola il guadagno medio e la deviazione standard.

[128 333; 19 720]

41 Una prova di tedesco contiene 80 difficoltà. La seguente tabella indica quanti errori sono stati commessi dagli studenti di una classe.

Errori	0	5	10	20	30	40	50
Studenti	2	1	5	4	4	3	2

Calcola il numero medio di errori e la deviazione standard. [22,62; 15,09]

42 Le misurazioni ripetute relative al consumo di energia elettrica per la prestazione giornaliera di una lavastoviglie in una mensa aziendale, presentano una distribuzione che si può ritenere gaussiana con media di 8 kWh e deviazione standard 0,85 kWh. Determina quante volte in 240 giorni il consumo è stato:

a) compreso fra 7,15 kWh e 8,85 kWh;
b) maggiore di 9,7 kWh;
c) minore di 5,45 kWh.

[a) 163,848; b) 5,46; c) 0,312]