

La statistica descrittiva

Considera un insieme di dati e li elabora:

- 1 presenta i dati in forma sintetica, grafica e/o tabulare;
- 2 caratterizza alcuni aspetti in modo sintetico: indici di posizione (es. valore medio), di dispersione (es. varianza), e di forma (es. simmetria);
- 3 studia le relazioni tra i dati riguardanti variabili diverse.

Tipi di variabili

I dati raccolti rappresentano la realizzazione (=valori che il caso ha pescato nell'esperimento) di variabili aleatorie.

Discrete e continue

Distinguiamo le variabili fra discrete e continue (vi ricordate la differenza?).

Ad esempio: le misure dell'apertura alare degli individui di una popolazione di rondini sono variabili continue; le loro età in anni sono variabili discrete.

Altro tipo di variabili

Noi ci occupiamo qui solo delle variabili numeriche, ma si possono trovare anche variabili non numeriche dette **categoriche** (es: il gruppo sanguigno).

Suddividiamo i dati in classi

Statistica
descrittivaSingole
variabili

Classi

Frequenze

Caso continuo

Istogrammi

Posizione

Dispersione

Forma

Indici per dati

raggruppati

Esempi

Variabili
multiple

Quando consideriamo una variabile, osservata su n individui, la lettura dei **dati grezzi** (= insieme di tutti i dati raccolti) può essere difficoltosa.

Per questo è utile **raggruppare** i dati in classi.

Ad esempio, supponiamo di avere raccolto dati su 200 spettatori di una certa trasmissione TV. In particolare una variabile osservata sia l'età.

Ci saranno molti valori uguali. Possiamo raggruppare in classi = intervalli di 5 anni, come nella tabella seguente.

Esempio delle età

Statistica
descrittivaSingole
variabili

Classi

Frequenze

Caso continuo

Istogrammi

Posizione

Dispersione

Forma

Indici per dati
raggruppati

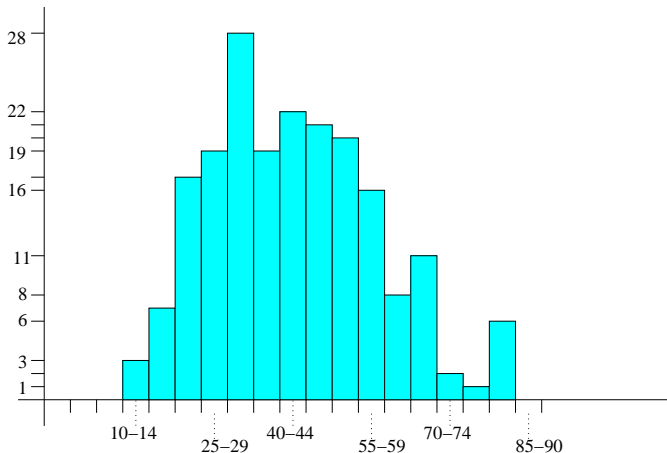
Esempi

Variabili
multiple

Cl.	Freq. Ass.	Freq. Rel.	Freq. Perc.	Freq. Perc. Cum.
10-14	3	0.015	1.5	1.5
15-19	7	0.035	3.5	5
20-24	17	0.085	8.5	13.5
25-29	19	0.095	9.5	23
30-34	28	0.14	14	37
35-39	19	0.095	9.5	46.5
40-44	22	0.11	11	57.5
45-49	21	0.105	10.5	68
50-54	20	0.1	10	78
55-59	16	0.08	8	86
60-64	8	0.04	4	90
65-69	11	0.055	5.5	95.5
70-74	2	0.01	1	96.5
75-79	1	0.005	0.5	97
80-84	6	0.03	3	100
85-90	0	0	0	100

Esempio delle età

Rappresentiamo la frequenza assoluta in un istogramma:



Frequenze

Una volta raggruppati gli n dati in classi si definiscono le frequenze. Le classi sono intervalli contigui. Nell'esempio i valori possibili sono numeri interi e le classi sono: $[10,14]$, $[15,19]$, ..., $[85,90]$.

FREQUENZE ASSOLUTA, RELATIVA, PERCENTUALE E CUMULATIVA

- 1 La **frequenza assoluta** f_a di una classe è il numero di osservazioni che ricadono in quella classe.
- 2 La **frequenza relativa** f_r di una classe è la sua frequenza assoluta divisa per il numero totale di osservazioni.
- 3 La **frequenza percentuale** f_p di una classe è la sua frequenza relativa moltiplicata per 100.
- 4 La **frequenza cumulativa** (F_a , F_r , F_p) di una classe è la somma delle frequenze della classe stessa e di tutte quelle che la precedono.

Classi per le variabili discrete

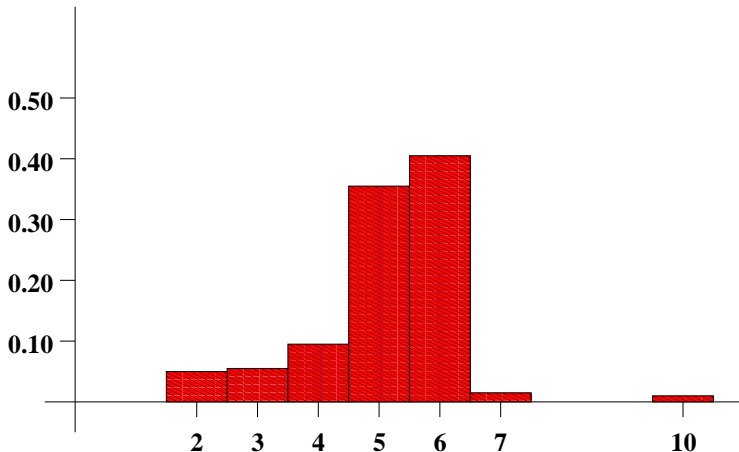
Se i valori diversi osservati non sono troppo numerosi, si può scegliere tutte le classi come singoli valori.

Esempio: le covate del *Passer Italiae*.

Numero di uova	Freq. assoluta	Freq. relativa
2	12	0.0522
3	15	0.0652
4	21	0.0913
5	82	0.3565
6	96	0.4174
7	3	0.0130
10	1	0.0044

Grafico per le covate

Rappresentiamo la frequenza relativa in un istogramma:



Caso continuo

Statistica
descrittivaSingole
variabili

Classi

Frequenze

Caso continuo

Istogrammi

Posizione

Dispersione

Forma

Indici per dati
raggruppati

Esempi

Variabili
multiple

Vediamo un esempio di variabili continue: la lunghezza di 50 petali di fiore di una specie di Iris:

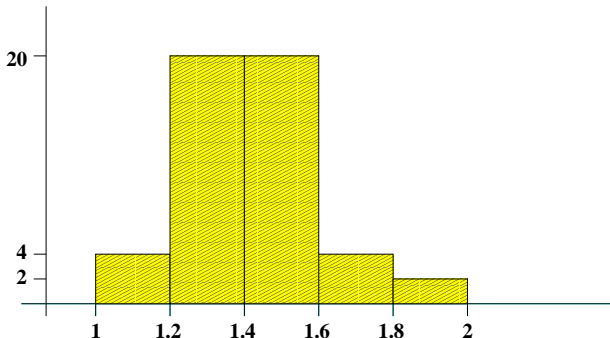
1.4, 1.4, 1.3, 1.5, 1.4, 1.7, 1.4, 1.5, 1.4, 1.5
1.5, 1.6, 1.4, 1.1, 1.2, 1.5, 1.3, 1.4, 1.7, 1.5
1.7, 1.5, 1.0, 1.7, 1.9, 1.6, 1.6, 1.5, 1.4, 1.6
1.6, 1.5, 1.5, 1.4, 1.5, 1.2, 1.3, 1.4, 1.3, 1.5
1.3, 1.3, 1.3, 1.6, 1.9, 1.4, 1.6, 1.4, 1.5, 1.4

Anche qui conviene raggruppare in intervalli, tenendo conto che i valori possibili sono (almeno) tutti i numeri reali fra 1 e 2 e che non si può inserire lo stesso dato in due classi (quindi le classi devono essere disgiunte).

Soluzione 1

Cl.	Freq. Ass.	Freq. Rel.	Freq. Perc.	Freq. Perc. Cum.
[1,1.2]	4	0.08	8	8
(1.2,1.4]	20	0.40	40	48
(1.4,1.6]	20	0.40	40	88
(1.6,1.8]	4	0.08	8	96
(1.8,2]	2	0.04	4	100

Rappresentiamo la frequenza assoluta in un istogramma:



Soluzione 2

Statistica
descrittivaSingole
variabili

Classi

Frequenze

Caso continuo

Istogrammi

Posizione

Dispersione

Forma

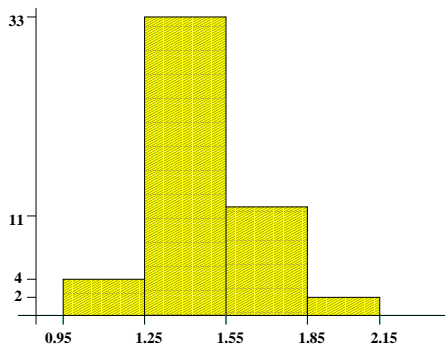
Indici per dati
raggruppati

Esempi

Variabili
multiple

Cl.	Freq. Ass.	Freq. Rel.	Freq. Perc.	Freq. Perc. Cum.
(0.95,1.25]	4	0.08	8	8
(1.25,1.55]	33	0.66	66	74
(1.55,1.85]	11	0.22	22	96
(1.85,2.15]	2	0.04	4	100

Rappresentiamo la frequenza assoluta in un istogramma:



Istogrammi

Gli istogrammi sono grafici in cui è rappresentata una scelta fra le frequenze assoluta, relativa e percentuale. Per questo:

- 1 i dati vengono suddivisi in classi = intervalli reali adiacenti, disgiunti e di **uguale lunghezza**.

Caso discreto

Si scelgono comunque le classi come intervalli adiacenti, anche se in questo modo gli estremi non risultassero valori possibili. Ad esempio nel caso delle uova anche se abbiamo scritto "2", la base del rettangolo corrispondente era [1.5,2.5). In questo modo non ci sono "buchi" fra i rettangoli (salvo per valori possibili non osservati).

- 2 Si disegnano rettangoli aventi come base una classe e altezza la sua frequenza (assoluta se stiamo rappresentando l'assoluta, etc).

La scelta delle classi

Notiamo che la suddivisione in classi è arbitraria: troppe classi portano a un grafico poco significativo; troppo poche classi fanno perdere informazioni (dai dati raggruppati non è possibile ricostruire i dati grezzi).

Classi di ampiezze diverse

Noi vediamo solo il caso in cui le **classi** sono intervalli tutti **di uguale ampiezza**.

Si possono anche trattare **classi di ampiezza diversa**. In tal caso solitamente è l'**area del rettangolo** ad essere **proporzionale alla frequenza**.

Indici di posizione: media

Si abbia un insieme di n dati x_1, \dots, x_n .

DEFINIZIONE DI MEDIA

La **media** è il numero:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Esempio (da Bramanti, Es.11): i dati siano 1, 2, 2, 2, 3, 3, 4, 6, 7. La media è

$$\frac{1}{9}(1 + 2 + 2 + 2 + 3 + 3 + 4 + 6 + 7) = 3.33.$$

Media = caso particolare di valore atteso

La media di un insieme di dati coincide con il valore atteso della variabile X = valore scelto a caso (cioè con uguale probabilità) fra i dati.

Indici di posizione: media

Media = caso particolare di valore atteso

La media di un insieme di dati si calcola utilizzando lo stimatore *media campionaria*; è quindi la stima puntuale del valore atteso della variabile aleatoria che modella la misura del dato in questione.

Trasformazione affine di dati.

Abbiamo delle osservazioni $\{x_1, x_2, \dots, x_n\}$ di cui abbiamo calcolato il valor medio \bar{x} . Ci interessa conoscere la media dei *dati trasformati in maniera affine* $y_i = ax_i + b$. Risulta

$$\bar{y} = a\bar{x} + b.$$

Infatti

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = \frac{1}{n} \sum_{i=1}^n (ax_i + b) = b + a \frac{1}{n} \sum_{i=1}^n x_i = a\bar{x} + b$$

Indici di posizione: mediana

Statistica
descrittivaSingole
variabiliClassi
Frequenze
Caso continuo
Istogrammi
Posizione
Dispersione
Forma
Indici per dati
raggruppati
EsempiVariabili
multiple

DEFINIZIONE DI MEDIANA

Si dispongono i dati in ordine crescente. La **mediana** è il dato nella posizione centrale se n è dispari, oppure la media aritmetica dei due dati in posizione centrale, se n è pari.

Esempio (da Bramanti, Es.11): i dati siano 1, 2, 2, 2, 3, 3, 4, 6, 7. I dati sono 9, quindi la mediana è il quinto dato, ovvero 3.

Esempio: i dati siano 1, 2, 2, 2, 3, 5, 5, 6, 7, 10. I dati sono 10, quindi la mediana è la media aritmetica del quinto dato (il 3) e del sesto dato (il 5), ovvero 4.

Indici di posizione: quantili, percentili, quartili

Sia $\{x_i\}_{i=1}^n$ un campione di dati numerici *ordinato* in maniera *non decrescente* (i.e. $x_i \leq x_j$ per ogni $i \leq j$).

Idea

Generalizziamo il concetto di mediana cercando un valore q_p con la proprietà che almeno una frazione p dei dati sia non superiore a q_p ed almeno una frazione $1 - p$ sia non inferiore a q_p ($0 < p < 1$).

Indici di posizione: quantili, percentili, quartili

Statistica
descrittivaSingole
variabili

Classi

Frequenze

Caso continuo

Istogrammi

Posizione

Dispersione

Forma

Indici per dati

raggruppati

Esempi

Variabili
multiple

DEFINIZIONE QUANTILE (non interpolata)

Dato $0 < p < 1$, il p -esimo quantile q_p è:

se np non è intero e k è l'intero tale che $k < np < k + 1$
allora $q_p = x_{k+1}$.

se $np = k$ con k intero, allora $q_p = (x_k + x_{k+1})/2$.

Calcoliamo la frazione di dati non superiore (risp. non inferiore) a q_p :

$$\frac{\#\{i : x_i \leq q_p\}}{n} \geq \begin{cases} \frac{k+1}{n} \geq p & k < np < k + 1 \\ \frac{k}{n} = p & k = np \end{cases}$$

$$\frac{\#\{i : x_i \geq q_p\}}{n} \geq \frac{n - k}{n} = 1 - \frac{k}{n} \geq 1 - p.$$

Indici di posizione: quantili, percentili, quartili

Statistica
descrittivaSingole
variabiliClassi
Frequenze
Caso continuo
Istogrammi
Posizione
Dispersione
Forma
Indici per dati
raggruppati
EsempiVariabili
multiple

- Il *p-esimo quantile* viene anche detto *100p-esimo percentile*.
- Il *p-esimo quantile* o *100p-esimo percentile* forniscono un valore che risulta maggiore o uguale del 100p% dei dati del campione.
- Il 25°, 50° e 75° percentile vengono detti anche primo, secondo e terzo *quartile*, e indicati con Q_1 , Q_2 , Q_3 .
- Q_1 , Q_2 , Q_3 sono tre numeri che dividono l'insieme di osservazioni in 4 gruppi contenenti ciascuno circa un quarto dei dati.
- Il secondo quartile Q_2 coincide con la mediana.

Indici di posizione: moda

DEFINIZIONE DI MODA

La **moda** è il valore o, più in generale, la classe in corrispondenza del quale si ha la popolazione più numerosa.

Si tratta dunque del punto dove la frequenza è massima.

DEFINIZIONE DI DISTRIBUZIONE UNI/PLURIMODALE

Se vi è un solo punto dove la frequenza è massima, si dice che la distribuzione delle frequenze è **unimodale**; se vi è più di un massimo, si dice che la distribuzione delle frequenze è **plurimodale**

Esempio (da Bramanti, Es.11): i dati siano 1, 2, 2, 2, 3, 3, 4, 6, 7. La moda è 2 (frequenza massima) e la distribuzione è unimodale.

Indici di dispersione: range

DEFINIZIONE DI RANGE

Se i dati sono x_1, x_2, \dots, x_n il range è il numero reale

$$r = \max\{x_i : i = 1, \dots\} - \min\{x_i : i = 1, \dots\}.$$

Esempio (da Bramanti, Es.11): i dati siano 1, 2, 2, 2, 3, 3, 4, 6, 7. Il range è $7-1=6$.

Indici di dispersione: varianza

DEFINIZIONE DI VARIANZA (di un insieme dati)

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \left(\frac{1}{n-1} \sum_{i=1}^n (x_i)^2 \right) - \frac{n}{n-1} (\bar{x})^2$$

Indici di dispersione: varianza

Statistica
descrittivaSingole
variabili

Classi

Frequenze

Caso continuo

Istogrammi

Posizione

Dispersione

Forma

Indici per dati

raggruppati

Esempi

Variabili
multiple

DEFINIZIONE ALTERNATIVA DI VARIANZA (di un insieme di dati)

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \left(\frac{1}{n} \sum_{i=1}^n (x_i)^2 \right) - (\bar{x})^2$$

Varianza di un insieme di dati = caso particolare di varianza di v.a.

La varianza di un insieme di dati coincide con la varianza della variabile X = valore scelto a caso (cioè con uguale probabilità) fra i dati.

Esempio (da Bramanti, Es.11): i dati siano 1, 2, 2, 2, 3, 3, 4, 6, 7. La varianza è $\frac{1}{9}(1 + 3 \cdot 4 + 2 \cdot 9 + 16 + 36 + 49) - (3.33)^2 = 3.56$.

Indici di dispersione: varianza

Statistica
descrittivaSingole
variabili

Classi

Frequenze

Caso continuo

Istogrammi

Posizione

Dispersione

Forma

Indici per dati

raggruppati

Esempi

Variabili
multiple

Trasformazione affine di dati.

Abbiamo delle osservazioni $\{x_1, x_2, \dots, x_n\}$ di cui abbiamo calcolato la varianza s_x^2 . Ci interessa conoscere la varianza dei *dati trasformati in maniera affine* $y_i = ax_i + b$. Risulta

$$s_y^2 = a^2 s_x^2.$$

Infatti

$$\begin{aligned} s_y^2 &= \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{1}{n-1} \sum_{i=1}^n (ax_i - a\bar{x})^2 \\ &= a^2 \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = a^2 s_x^2. \end{aligned}$$

Standardizzazione di un campione

Dato il campione $\{x_1, \dots, x_n\}$, definiamo la sua *standardizzazione* operando la seguente trasformazione

$$y = \frac{x - \bar{x}}{s_x}$$

Il *campione standardizzato* corrispondente $y_i = (x_i - \bar{x})/s_x$ ha media nulla e varianza 1. Infatti

$$\bar{y} = \frac{1}{s_x}(\bar{x} - \bar{x}) = 0$$

$$s_y^2 = \frac{1}{s_x^2} s_x^2 = 1.$$

Indici di forma: skewness

DEFINIZIONE DI SKEWNESS (o COEFFICIENTE DI ASIMMETRIA)

La **skewness**

$$\gamma_3 = \frac{1}{n\sigma^3} \sum_{i=1}^n (x_i - \bar{x})^3,$$

dove σ è la radice della varianza.

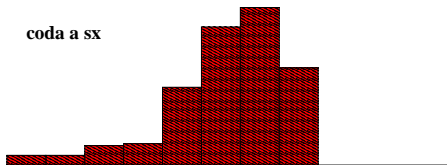
A colpo d'occhio.

Se è negativa denota una *coda* verso sinistra.

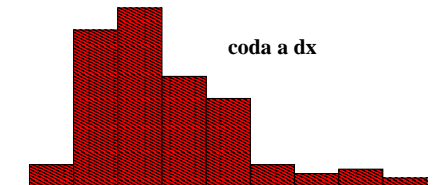
Se è positiva denota una *coda* verso destra.

Se la distribuzione è simmetrica, allora la skewness è nulla, ma l'inverso non è vero.

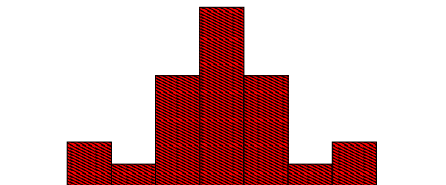
Esempio di distribuzione con skewness negativa:



Esempio di distribuzione con skewness positiva:



Esempio di distribuzione con skewness = 0:



Indici di forma: curtosi

DEFINIZIONE DI CURTOSI (O KURTOSIS)

La **curtosi**

$$\gamma_4 = \frac{1}{n\sigma^4} \sum_{i=1}^n (x_i - \bar{x})^4,$$

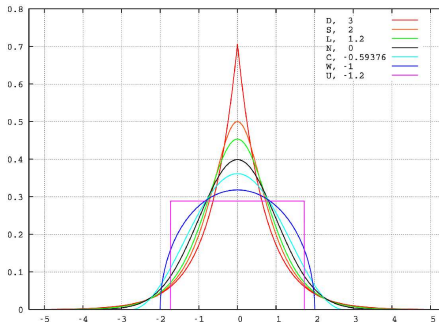
dove σ è la radice della varianza.

Proprietà.

È un numero ≥ 0 . Misura (in un certo senso) quanto è “appuntita” la distribuzione delle frequenze.

Valori elevati della curtosi segnalano distribuzioni “piccate”, valori piccoli si hanno generalmente in corrispondenza di distribuzioni meno appuntite.

Solitamente si confronta con la $\mathcal{N}(0, 1)$ che ha curtosi = 3. In questo grafico (da Wikipedia) diverse distribuzioni (tutte con skewness = 0) e i rispettivi valori di $\gamma_4 - 3$.



Unità di misura degli indici

Unità di misura degli indici

- 1 La media, la mediana, la moda e il range hanno la stessa unità di misura dei dati.
- 2 La varianza ha l'unità di misura dei dati al quadrato.
- 3 Skewness e curtosi sono numeri puri.

Media per dati raggruppati in classi

Statistica
descrittivaSingole
variabili

Classi

Frequenze

Caso continuo

Istogrammi

Posizione

Dispersione

Forma

Indici per dati
raggruppati

Esempi

Variabili
multiple

Dividiamo i dati in N_c classi indicando con x_{kl} il dato l -esimo della k -esima classe e con $f_a(k)$ la frequenza assoluta della k -esima classe. Possiamo riorganizzare il calcolo della media nel seguente modo

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} \sum_{k=1}^{N_c} \sum_{l=1}^{f_a(k)} x_{kl}$$

ma, per definizione,

$$\sum_{l=1}^{f_a(k)} x_{kl} = f_a(k) \bar{x}_k$$

se \bar{x}_k è la media dei dati della classe k -esima.

Media per dati raggruppati in classi

Statistica
descrittivaSingole
variabili

Classi

Frequenze

Caso continuo

Istogrammi

Posizione

Dispersione

Forma

Indici per dati
raggruppati

Esempi

Variabili
multiple

Sostituendo si ha:

$$\bar{x} = \frac{1}{n} \sum_{k=1}^{N_c} f_a(k) \bar{x}_k = \sum_{k=1}^{N_c} f_r(k) \bar{x}_k$$

Osservazione

La media si ottiene dalle frequenze assolute o relative delle classi dei dati raggruppati se sono noti i valori medi dei dati in ciascuna classe.

Se non sono noti, si sostituisce a ciascun \bar{x}_k il valore centrale dell'intervallo associato alla classe k .

In tal modo si ottiene un valore approssimato della media.

Varianza per dati raggruppati in classi

Ricordiamo la formula della varianza

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \sum_{k=1}^{N_c} \sum_{l=1}^{f_a(k)} (x_{lk} - \bar{x})^2$$

sostituiamo x_{lk} con \bar{x}_k

$$\begin{aligned} s^2 &\approx \frac{1}{n-1} \sum_{k=1}^{N_c} \sum_{l=1}^{f_a(k)} (\bar{x}_k - \bar{x})^2 \\ &= \frac{1}{n-1} \sum_{k=1}^{N_c} f_a(k) (\bar{x}_k - \bar{x})^2 = \frac{1}{n-1} \sum_{k=1}^{N_c} f_a(k) \bar{x}_k^2 - \frac{n}{n-1} \bar{x}^2 \\ &= \frac{n}{n-1} \sum_{k=1}^{N_c} f_r(k) (\bar{x}_k^2 - \bar{x}^2) = \frac{n}{n-1} \left(\sum_{k=1}^{N_c} f_r(k) \bar{x}_k^2 - \bar{x}^2 \right). \end{aligned}$$

Percentili per dati raggruppati in classi

In presenza di dati raggruppati si può stimare il quantile q_α individuando la classe cui appartiene.

La classe in cui cade sarà quella in corrispondenza alla quale si supera il valore α nel calcolo della frequenza relativa cumulativa, cioè

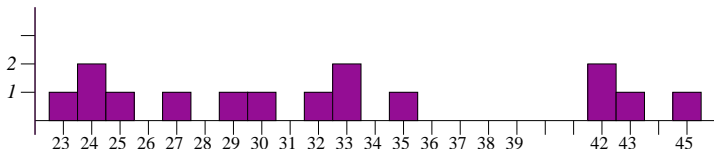
$$q_\alpha \in I_k \text{ dove } k := \min\{i = 1, \dots, N_c : F_r(i) \geq \alpha\}.$$

Osservazione

Esiste anche una stima interpolata (si vedano le dispense).

Rivediamo gli insiemi di dati visti in precedenza.
Cominciamo con i dati relativi alle uova deposte da una pulce in 15 giorni diversi:

24, 35, 45, 43, 25, 33, 33, 30, 29, 27, 32, 24, 23, 42, 42.



Media = 32.47

Mediana = 32

Moda = 24 e 42

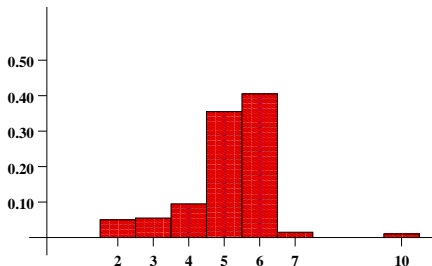
Range = 22

Varianza = 52.92

Skewness = 0.43

Curtosi = 1.83.

Questo è un esempio di istogramma poco significativo. Sarebbe stato meglio raggruppare i dati in classi più ampie.



Media = 5.09

Mediana = 5

Moda = 6

Range = 8

Varianza = 1.39

Skewness = -0.81

Curtosi = 4.76.

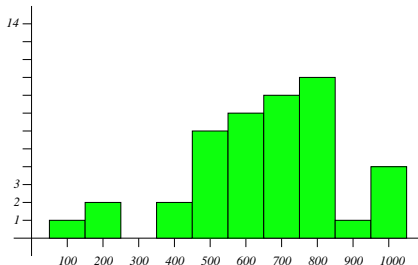
Il primo paese

Statistica
descrittiva

Singole
variabili

Classi
Frequenze
Caso continuo
Istogrammi
Posizione
Dispersione
Forma
Indici per dati
raggruppati
Esempi

Variabili
multiple



Media = 655

Mediana = 700

Moda = 800

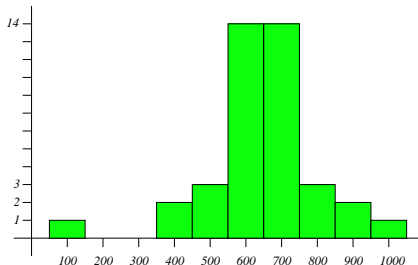
Range = 900

Varianza = 43975

Skewness = -0.59

Curtosi = 3.53.

Il secondo paese



Media = 645

Range = 900

Skewness = -0.88

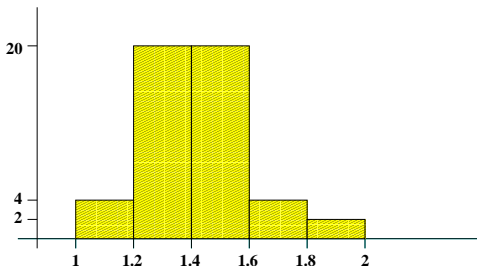
Mediana = 650

Varianza = 21975

Curtosi = 7.00

Moda = 600 e 700

I petali di Iris

Statistica
descrittivaSingole
variabiliClassi
Frequenze
Caso continuo
Istogrammi
Posizione
Dispersione
Forma
Indici per dati
raggruppati
EsempiVariabili
multiple

Media = 1.46

Mediana = 1.5

Moda = 1.4 e 1.5

Range = 0.9

Varianza = 0.023

Skewness = 0.1064

Curtosi = 4.02

La moda, se guardassimo anziché i valori le classi, è costituita dalle due classi $(1.2, 1.4]$ e $(1.4, 1.6]$. (Con l'altra suddivisione in classi avremmo invece una sola moda: $(1.25, 1.55]$).

Caso bidimensionale

Per semplicità ci occupiamo solo del caso bidimensionale (il caso multidimensionale è analogo e si trova sulle dispense).

- Consideriamo coppie di dati $\{(x_i, y_i)\}_{i=1}^n$.
- Suddividiamo i dati in N_x e N_y classi disgiunte rispetto alla prima e seconda coordinata, siano esse $\{C_x(i)\}_{i=1}^{N_x}$ e $\{C_y(j)\}_{j=1}^{N_y}$.

Caso bidimensionale

Le frequenze assolute e relative congiunte sono quindi

$$f_a(i, j) := \#C_x(i) \cap C_y(j), \quad f_r(i, j) := f_a(i, j)/n$$

per ogni $i = 1, \dots, N_x, j = 1, \dots, N_y$.

Le frequenze cumulative assolute e relative congiunte sono
è invece

$$F_a(i, j) := \sum_{t \leq i, h \leq j} f_a(t, h),$$

$$F_r(i, j) := F_a(i, j)/n \equiv \sum_{t \leq i, h \leq j} f_r(t, h).$$

per ogni $i = 1, \dots, N_x, j = 1, \dots, N_y$.

Caso bidimensionale

Si vede facilmente che

$$f_a^x(i) = \sum_{j=1}^{N_y} f_a(i, j) \quad f_a^y(j) = \sum_{i=1}^{N_x} f_a(i, j)$$

rappresentano le frequenze monodimensionali assolute dei campioni $\{x_i\}_{i=1}^n$ e $\{y_i\}_{i=1}^n$ rispettivamente.

f_a^x e f_a^y prendono il nome di *frequenze marginali* assolute.

Analogamente si procede per le frequenze relative.