

Esercizi per il corso di Probabilità e Statistica

Foglio 0: Statistica descrittiva

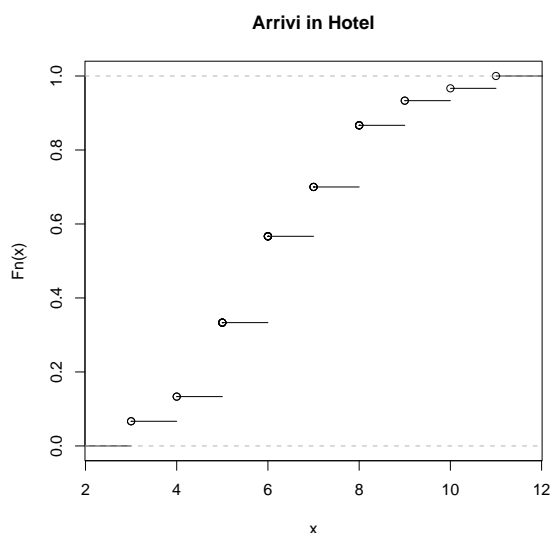
- Dati i seguenti valori di X : $x_1 = 3$, $x_2 = 1$, $x_3 = 4$, $x_4 = 6$, $x_5 = 5$,
 - calcolare $\sum_{i=1}^5 x_i$, $\prod_{i=1}^5 x_i$ e $\sum_{i=1}^5 \sum_{j=1}^i x_j$;
 - dimostrare che $(\sum_{i=1}^n x_i)^2 = \sum_{i=1}^n x_i^2 + 2 \sum_{i=1}^n \sum_{j>i}^n x_i x_j$;
 - dimostrare che la media \bar{x} minimizza la funzione $f(a) = \sum_{i=1}^n (x_i - a)^2$.
- Per ognuna delle seguenti variabili si dica se si tratta di una variabile quantitativa (discreta o continua) o qualitativa (nominale o ordinale):
 - l'età degli studenti in questo corso;
 - i programmi preferiti dalle studentesse di questo corso;
 - il numero degli studenti di questo corso che guardano la trasmissione "Il Grande Fratello";
 - le tasse universitarie pagate dagli studenti di questo corso;
 - il voto riportato all'esame di matematica;
 - l'aver superato l'esame di matematica.
- Il docente di un corso universitario ha raccolto alcuni dati sugli studenti che hanno sostenuto un particolare esame. La tabella riporta i dati relativi ai primi quattro studenti.

Voto	Anno di corso	Residenza	Diploma Sup.	Frequentante
28	I	Treviso	Liceo	S
21	IV	Venezia	Ist. Tecnico	N
18	I	Verona	Ist. Comm.	S
21	II	Padova	Ist. Prof.	S

- Qual è l'unità statistica? Qual è la popolazione di riferimento?
 - Quali caratteri (variabili) sono rilevati?
 - Quali sono le modalità rilevate del Diploma Superiore?
 - Quali sono le modalità rilevate del Voto e quali quelle possibili?
- In un gruppo di 20 persone sono state rilevate due variabili, il sesso e l'età

47	61	38	40	26	41	49	65	53	55
F	M	M	M	F	M	F	F	M	F
30	23	34	33	40	21	65	32	47	50
M	F	M	F	M	M	M	F	F	F

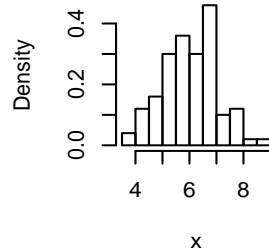
- (a) Si costruisca la distribuzione per classi dell'età (utilizzando le classi 19-29, 30-44, 45-59, 60 e oltre) e si calcolino le frequenze assolute, relative, percentuali, le cumulate, e le densità di frequenza. Si disegni un opportuno istogramma.
 - (b) Si rappresenti la funzione di ripartizione empirica e se ne calcoli il valore nel punto 57.
 - (c) Si disegni il diagramma a scatola con baffi, partendo dai dati originari.
 - (d) Si calcolino media, varianza, scarto quadratico medio, campo di variazione, scarto interquantile e coefficiente di variazione.
 - (e) Si chiamino y_i i valori originari e z_i quelli standardizzati: si trovino media e varianza degli z_i attraverso le corrispondenti quantità degli y_i .
5. In questo grafico mostriamo la funzione di ripartizione empirica degli arrivi giornalieri in un hotel rilevata nel mese di aprile (il cerchietto indica che il punto è incluso).



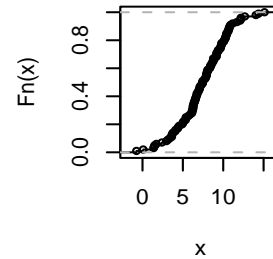
Si dica quali delle seguenti affermazioni sono vere:

- (a) il 50% degli arrivi giornalieri è inferiore o uguale a 6;
 - (b) il massimo delle osservazioni è superiore o uguale a 11 arrivi;
 - (c) il minimo delle osservazioni è superiore a 3 arrivi.
6. Si associ ad ogni istogramma delle densità di frequenza la corrispondente funzione di ripartizione empirica.

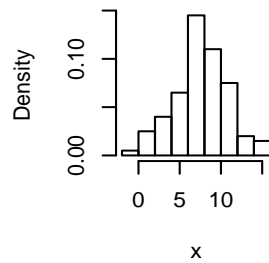
Istogramma 1



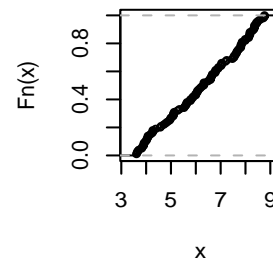
Funzione di ripartizione A



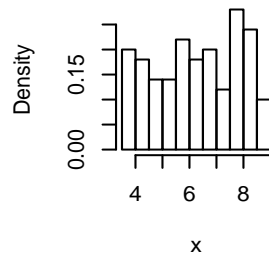
Istogramma 2



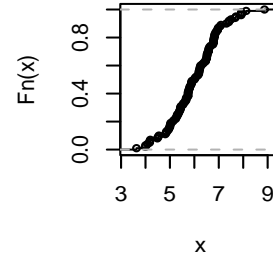
Funzione di ripartizione B



Istogramma 3



Funzione di ripartizione C



7. Si è rilevato il numero di clienti giornalieri di un negozio per 14 giorni:

4, 10, 3, 15, 15, 7, 4, 7, 9, 25, 6, 4, 8, 8.

- (a) Si calcolino le frequenze assolute, relative e cumulate;
- (b) si calcolino la media, la mediana, il primo e il terzo quartile, lo scarto interquartile;
- (c) si disegni il diagramma a scatola con baffi.

8. È riportata di seguito una tabella che riassume le lunghezze di 100 foglie di platano, misurate al millimetro più prossimo, dopo 10 giorni di siccità.

Lunghezza	Freq.	Freq. cumulata
120 + 135	10	10
135 + 145	20	30
145 + 150	60	90
150 + 165	10	100

A partire dalla tabella di frequenza

- (a) si disegni per queste osservazioni un opportuno istogramma;
 - (b) si calcolino mediana, primo quartile, terzo quartile, scarto interquartile e media.
 - (c) È noto che l'effetto di un giorno di pioggia continuata è di aumentare la lunghezza delle foglie di una quota percentuale pari al 10% più una quota fissa pari a 0.5 mm. Si calcolino mediana, scarto interquartile e media dopo un giorno di pioggia continuata.
9. Una fabbrica di televisori produce due tipi di tubi catodici: il tipo A e il tipo B. I tubi catodici hanno tempi di durata media, rispettivamente di 1495 ore e 1875 ore e scarti quadratici medi rispettivamente di 280 ore e 310 ore. In generale, è preferibile il tubo catodico che ha la durata più alta e variabilità più bassa. Quale dei due tubi catodici è preferibile?
10. I seguenti dati sono relativi ai minuti di ritardo con cui un treno è arrivato a destinazione negli ultimi dieci giorni (un numero negativo sta ad indicare che il treno è arrivato in anticipo per i minuti corrispondenti):

$$-3, 6, 4, 10, -4, 124, 2, -1, 4, 1.$$

- (a) Se foste assunti dalle ferrovie per mostrare che le ferrovie forniscono un buon servizio, quali misure di posizione usereste?
- (b) Se foste assunti da una televisione locale per mostrare che le ferrovie forniscono un cattivo servizio, quali misure di posizione usereste?
- (c) Se cercaste di essere obiettivi e imparziali nel valutare le ferrovie, quali misure di posizione usereste?
- (d) Calcolate il campo di variazione, la varianza, lo scarto quadratico medio e il coefficiente di variazione per il "ritardo" (in minuti).
- (e) Commentate la variabilità dei dati.
- (f) Quale sarebbe l'effetto sulle vostre conclusioni in (a)-(e) se al posto di 124 si fosse scorrettamente rilevato 12?
- (g) Descrivete la forma della distribuzione dei dati originari.
- (h) Descrivete la forma della distribuzione dei dati se a 124 è sostituito 12.
- (i) Calcolate la funzione di ripartizione empirica e dite se è vero che: $F(0) = 0$, $F(124) = 1$, $F(140) = 1$.

11. In un negozio di abbigliamento si è registrato quali capi venivano venduti in un particolare giorno. Il risultato è stato:

$P, P, M, G, P, M, M, G, P, G, G, M, P, M, M, G, M,$

dove P =pantalone, G =giacca, M =maglia.

- (a) Calcolare la distribuzione delle frequenze assolute e relative; disegnare il diagramma a barre;
 - (b) identificare la moda;
 - (c) calcolare l'indice di Gini normalizzato e l'entropia di Shannon normalizzata.
12. La seguente tabella si riferisce alla distribuzione della popolazione residente attiva (in migliaia di unità) per settore di attività (Censimento generale della popolazione italiana, 1961 e 1971). In riferimento sia al 1961 sia al 1971:

m_i	1961 f_i	1971 f_i
Agricoltura, foreste, caccia, pesca	5693	3242
Industrie	7963	8350
Altre attività	5936	7238

- (a) Disegnare i relativi diagrammi a barre e identificare le relative mode.
 - (b) Calcolare l'indice di Gini normalizzato e l'indice di Shannon normalizzato. Commentare.
13. In un sondaggio riguardante un certo problema politico, si vuole sapere se i votanti Conservatori vedano il problema diversamente da quelli Laburisti. Si considerano 250 individui, ottenendo i seguenti risultati:

Partito/Opinione	Contrari	Incerti	Favorevoli	Totale
Conservatori	40	10	50	100
Laburisti	60	40	50	150
Totale	100	50	100	250

- (a) Si calcolino le distribuzioni marginali e condizionate e le si rappresentino graficamente.
 - (b) Si verifichi se la differenza di opinione non è dovuta alla preferenza politica.
14. Per verificare l'effetto di un farmaco anti-emicrania, 100 pazienti sono stati trattati con il farmaco sotto studio mentre 200 sono stati trattati con il farmaco in uso al momento dello studio. Poi per ciascun paziente è stata raccolta la risposta codificata come nessun effetto, dolore diminuito, dolore assente. I risultati sono sintetizzati nella seguente tabella:

farmaco	nessun effetto	dolore diminuito	dolore assente
nuovo	41	36	23
vecchio	76	81	43

- (a) Rappresentare graficamente i dati.
- (b) Analizzare la tabella con gli strumenti a voi noti.

15. Nella tabella sottostante sono riportati i consumi (in milioni di tonnellate) di risorse naturali impiegate nell'economia italiana nel quinquennio 2000-2004 e le importazioni delle medesime risorse nello stesso arco temporale.

	2000	2001	2002	2003	2004
Fabbisogno	2.357	2.295	2.214	2.077	2.184
Importazioni	329.028	330.035	334.807	343.784	360.282

- (a) Si rappresentino graficamente i dati;
- (b) si calcolino la covarianza e il coefficiente di correlazione e si dia un'interpretazione dei risultati.

16. Nella tabella sottostante sono riportati i dati riguardanti lo spessore dell'ossidazione di una lega metallica posta a diverse temperature in un forno.

	1	2	3	4	5
Ossidazione	4.00	3.00	3.00	2.00	2.00
Temperatura	-2.00	-1.00	0.00	1.00	2.00

- (a) Si rappresentino graficamente i dati;
- (b) si calcolino la covarianza e il coefficiente di correlazione e si dia un'interpretazione dei risultati.