Statistica e Laboratorio: esercizi

Paolo Vidoni
Dipartimento di Scienze Economiche e Statistiche, Università di Udine

Settembre 2021

A) Statistica descrittiva

A.1) Nella seguente tabella sono riportate le lunghezze di 40 foglie di platano, registrate al millimetro più prossimo

138	164	150	132	144	125	149	157
146	158	140	147	136	148	152	144
168	126	138	176	163	119	154	165
146	173	142	147	135	153	140	135
161	145	135	142	150	156	145	128

Si definisca la variabile Y, oggetto di interesse, e si dica se è qualitativa, sconnessa o ordinale, o quantitativa, discreta o continua. Con riferimento ai risultati della rilevazione, si specifichi la associata variabile statistica e si individui il suo supporto. Dopo aver definito le seguenti classi di modalità di ampiezza 10 mm, $118 \dashv 128$, $128 \dashv 138, \cdots$, $168 \dashv 178$, si specifichi la distribuzione di frequenza assoluta e relativa e le corrispondenti frequenze cumulate. Si costruiscano le rappresentazioni grafiche mediante l'istogramma e il poligono di frequenza. Inoltre, si determini la moda, la mediana e la media aritmetica di Y e si esprima un parere sulla simmetria della distribuzione di frequenza. Si calcolino i quantili di livello $\alpha = 0.25$ e $\alpha = 0.75$ e si costruisca la rappresentazione grafica mediante il diagramma a scatola e baffi. Infine, si calcoli il campo di variazione, lo scarto interquartilico, la varianza e il coefficiente di variazione.

A.2) Si sono rilevati i guasti meccanici riportati da 40 autovetture FIAT e da 60 autovetture OPEL, possedute da una azienda di spedizioni, nei primi 50.000 Km di percorrenza. I dati sono sintetizzati nella seguente tabella che riporta le frequenze assolute

No. guasti	Autovetture FIAT	Autovetture OPEL
0	9	33
1	13	20
2	10	6
3	5	1
4	3	0
Totale	40	60

Si definisca la variabile Y, oggetto di interesse, e si dica se è qualitativa, sconnessa o ordinale, o quantitativa, discreta o continua. Con riferimento ai risultati della rilevazione, sintetizzati nella tabella di frequenza, si specifichino le due variabili statistiche ed i rispettivi supporti. Si determinino le distribuzioni di frequenza relativa e si costruiscano le rappresentazioni grafiche mediante diagrammi con bastoncini. Si calcolino le frequenze cumulate, relative e assolute, delle due variabili statistiche. Inoltre, si determinino la moda, la mediana e la media aritmetica e si calcolino la varianza e il coefficiente di variazione. Si esprima un parere sul confronto tra le due popolazioni di autovetture.

A.3) Le votazioni finali in centesimi riportate da 80 studenti in un test di ammissione al Corso di Laurea in Biotecnologie, sono registrate nella seguente tabella

68	84	75	82	68	90	62	88	76	93
73	79	88	73	60	93	71	59	85	75
61	65	75	87	74	62	95	78	63	72
66	78	82	75	94	77	69	74	68	60
96	78	89	61	75	95	60	79	83	71
79	62	67	97	78	85	76	65	71	75
65	80	73	57	88	78	62	76	53	74
86	67	73	81	72	63	76	75	85	77

Si definisca la variabile Y, oggetto di interesse, e si dica se è qualitativa, sconnessa o ordinale, o quantitativa, discreta o continua. Con riferimento ai risultati della rilevazione, si specifichi la associata variabile statistica e si individui il suo supporto. Si specifichi la distribuzione di frequenza assoluta e relativa e le corrispondenti frequenze cumulate, introducendo, se risulta opportuno, un conveniente raggruppamento delle modalità contigue in classi. Si calcoli la frequenza relativa di studenti che hanno riportato una votazione più alta di 65 e inferiore a 85. Si proponga una utile rappresentazione grafica per la distribuzione di frequenza relativa. Inoltre, si determini la moda, la mediana e la media aritmetica di Y, si calcolino i quantili di livello $\alpha = 0.25$ e $\alpha = 0.75$ e si costruisca la rappresentazione grafica mediante il diagramma a scatola e baffi. Infine, si calcoli il campo di variazione, lo scarto interquartilico, la varianza e il coefficiente di variazione.

A.4) La distribuzione di frequenza assoluta degli stipendi mensili in euro dei 65 dipendenti di una determinata azienda è riportata nella seguente tabella

Stipendio	No. di dipendenti
$1000 \dashv 1100$	8
$1100 \dashv 1200$	10
$1200 \dashv 1300$	16
$1300 \dashv 1400$	14
$1400 \dashv 1500$	10
$1500 \dashv 1800$	5
$1800 \dashv 2500$	2
Totale	65

Si definisca la variabile Y, oggetto di interesse, e si dica se è qualitativa, sconnessa o ordinale, o quantitativa, discreta o continua. Con riferimento ai risultati della rilevazione, sintetizzati nella tabella di frequenza, si specifichi la associata distribuzione di frequenza relativa e le corrispondenti frequenze cumulate, assolute e relative. Si costruiscano le rappresentazioni grafiche mediante l'istogramma e il poligono di frequenza. Infine, si individui la classe mediana e, dopo aver fatto convenienti assunzioni semplificative, si calcoli la media aritmetica e la varianza.

A.5) Su 15 studenti iscritti al secondo anno presso la Facoltà di Lettere e Filosofia dell'Università di Udine si sono rilevate le seguenti variabili: provincia di residenza (con modalità, UD=Udine, AL=altro), corso di laurea (con modalità, C=Conservazione dei beni culturali, L=Lettere, D=DAMS) e grado di soddisfazione (con modalità, 1=insoddisfatto, 2=poco soddisfatto, 3=soddisfatto, 4=molto soddisfatto). I risultati della rilevazione sono riportati nella seguente matrice dei dati

Studente	Provincia	Corso	Soddisfazione
1	UD	C	1
2	UD	C	1
3	UD	L	2
4	AL	D	3
5	AL	D	3
6	AL	L	2
7	UD	L	2
8	UD	D	4
9	UD	C	4
10	AL	C	3
11	AL	C	3
12	UD	L	4
13	AL	D	2
14	AL	C	1
15	AL	D	3

Si individuino le variabili di interesse e si dica se sono qualitative, sconnesse o ordinali, o quantitative, discrete o continue. Con riferimento ai risultati della rilevazione, si specifichino le associate variabili statistiche, i loro supporti e le loro distribuzioni di frequenza assolute e relative. Si proponga una conveniente rappresentazione grafica per le tre serie statistiche e si determinino, ove hanno significato, le frequenze cumulate, assolute e relative. Infine, con riferimento alle tre variabili statistiche, si costruiscano degli opportuni indici di posizione.

A.6) La distribuzione di frequenza assoluta del peso di 100 studenti maschi è riportata nella seguente tabella, con i valori espressi in chilogrammi,

Peso	No. studenti
$60 \dashv 62$	5
$62 \dashv 65$	18
$65 \dashv 68$	42
$68 \dashv 71$	27
$71 \dashv 73$	8
Totale	100

Si definisca la variabile Y, oggetto di interesse, e si dica se è qualitativa, sconnessa o ordinale, o quantitativa, discreta o continua. Con riferimento ai risultati della rilevazione, sintetizzati nella tabella di frequenza, si specifichi la associata distribuzione di frequenza relativa e le corrispondenti frequenze cumulate, assolute e relative. Si costruisca una rappresentazione grafica mediante l'istogramma. Inoltre, si individui la classe mediana e, dopo aver fatto convenienti assunzioni semplificative, si calcoli la media aritmetica e la varianza di Y.

A.7) La distribuzione di frequenza assoluta dell'età, in anni compiuti (a.c.), dei laureati in Italia nel 1986 per i corsi di laurea in Ingegneria Elettronica e Lettere è riportata nella seguente tabella, con i valori raggruppati in classi,

Età in a.c.	No. laureati Ingegneria El.	No. laureati Lettere
$23 \vdash 25$	9	282
$25 \vdash 27$	405	1432
$27 \vdash 29$	710	1065
$29 \vdash 31$	252	285
$31 \vdash 40$	237	591
Totale	1613	3655

Si definisca la variabile Y, oggetto di interesse, e si dica se è qualitativa, sconnessa o ordinale, o quantitativa, discreta o continua. Con riferimento ai risultati della rilevazione, si specifichi per entrambi i corsi di laurea la associata distribuzione di frequenza relativa e le corrispondenti frequenze cumulate, assolute e relative. Si costruisca una opportuna rappresentazione grafica, si individui la classe mediana, la media aritmetica e la varianza. Infine, si confrontino le due distribuzioni di frequenza e si faccia un breve commento.

A.8) Sono riportate di seguito le spese per l'elettricità in euro, relative al mese di luglio 2002, rilevate su 50 appartamenti in una zona residenziale di Milano:

96	171	202	178	147	102	153	197	127	82	157	185	90	116	172	111
148	213	130	165	141	149	206	175	123	128	144	168	109	167	95	163
150	154	130	143	187	166	139	149	108	119	183	151	114	135	191	137
129	158														

Si definisca la variabile Y, oggetto di interesse, e si dica se è qualitativa, sconnessa o ordinale, o quantitativa, discreta o continua. Si specifichino le associate distribuzione di frequenza assoluta e relativa (non cumulata e cumulata), sulla base della seguente ripartizione in classi: $70 \dashv 100, \cdots, 190 \dashv 220$.

Si costruisca una rappresentazione grafica mediante l'istogramma. Inoltre, si individui la classe mediana e, dopo aver fatto convenienti assunzioni semplificative, si calcoli la media aritmetica e la varianza di Y. Infine, dopo aver calcolato media, mediana e quartili sui dati grezzi, si esprima un parere sulla simmetria della distribuzione di frequenza e si disegni il diagramma a scatola e baffi.

A.9) Si consideri la seguente tabella che fornisce informazioni sul numero di figli per nucleo famigliare, con riferimento a 100 famiglie di un determinato quartiere

No. figli	0	1	2	3	4	5	Totale
No. famiglie	10	42	30	11	5	2	100

Si definisca la variabile Y, oggetto di interesse, e si dica se è qualitativa, sconnessa o ordinale, o quantitativa, discreta o continua. Con riferimento ai risultati della rilevazione, sintetizzati nella tabella di frequenza, si specifichi la associata distribuzione di frequenza relativa e le corrispondenti frequenze cumulate, assolute e relative. Si costruisca una opportuna rappresentazione grafica per la distribuzione di frequenza relativa di Y. Inoltre, si calcolino la media aritmetica, la moda, la mediana e la varianza di Y.

A.10) Si consideri la seguente tabella che fornisce la distribuzione di frequenza assoluta della concentrazione di ozono nell'atmosfera (in parti di miliardi) rilevata in 136 giorni in una città italiana

Concentr. ozono	$0 \vdash 50$	$50 \vdash 75$	$75 \vdash 100$	$100 \vdash 150$	$150 \vdash 200$	$200 \vdash 250$	Totale
Frequenza	35	29	25	28	11	8	136

Si definisca la variabile Y, oggetto di interesse, e si dica se è qualitativa, sconnessa o ordinale, o quantitativa, discreta o continua. Con riferimento ai risultati della rilevazione, sintetizzati nella tabella di frequenza, si specifichi la associata distribuzione di frequenza relativa e le corrispondenti frequenze cumulate, assolute e relative. Si costruisca una opportuna rappresentazione grafica per la distribuzione di frequenza relativa di Y. Inoltre, si individui la classe mediana e, dopo aver fatto convenienti assunzioni semplificatrici, si calcolino la media aritmetica e la varianza di Y.

A.11) Si consideri la seguente tabella che fornisce la distribuzione di frequenza assoluta del numero di interruzioni settimanali, con riferimento alle ultime 56 settimane, in una determinata linea di produzione di una grande azienda automobilistica

No. interruzioni	0	1	2	3	4	5	6	7	8	Totale
frequenza	6	14	10	8	7	5	3	2	1	56

Si definisca la variabile Y, oggetto di interesse, e si dica se è qualitativa, sconnessa o ordinale, o quantitativa, discreta o continua. Con riferimento ai risultati della rilevazione, sintetizzati nella tabella di frequenza, si specifichi la associata distribuzione di frequenza relativa e le corrispondenti frequenze cumulate, assolute e relative. Si costruisca una opportuna rappresentazione grafica per la distribuzione di frequenza relativa di Y. Inoltre, si individui la mediana e la moda e si calcolino la media aritmetica e la varianza di Y. Si esprima un parere sulla simmetria della distribuzione di frequenza.

A.12) Si consideri la seguente tabella che fornisce la distribuzione di frequenza assoluta dei laureati di cittadinanza estera in atenei italiani nel 2006, ripartita per macro-aree geografiche (Dati *AlmaLaurea*)

Provenienza	No. laureati
Unione Europea	1398
Europa extra UE	1609
Asia	349
Africa	332
Americhe	500
Oceania	12
Totale	4200

Si definisca la variabile Y, oggetto di interesse, e si dica se è qualitativa, sconnessa o ordinale, o quantitativa, discreta o continua. Con riferimento ai risultati della rilevazione, si specifichi la associata distribuzione di frequenza relativa e si costruisca una opportuna rappresentazione grafica.

A.13) Con riferimento ai clienti presenti in un dato negozio, sono state rilevate le età in anni compiuti. I risultati della rilevazione sono 16, 23, 40, 29, 16, 21. Si rappresentino tali dati con la funzione di ripartizione empirica.

A.14) L'osservazione di una variabile quantitativa discreta Y su una popolazione di $n = 56$ unità statistiche ha fornito la seguente tabella con le frequenze assolute

Y	freq. ass.
1	6
2	11
3	22
5	6
7	11
Totale	56

Con riferimento a tali risultati, si specifichi la associata distribuzione di frequenza relativa e le corrispondenti frequenze cumulate, assolute e relative. Si definisca e si disegni la funzione di ripartizione empirica.

A.15) Un'indagine condotta su $n = 21$ studenti fuori sede ha riscontrato che la cifra mensile per le spese di affitto in euro è

210 170 90 190 80 200 220 140 120 130 110 320 160 260 150 650
110 310 160 190 280

Si calcolino la media, la varianza e i quartili. Dopo aver definito le seguenti classi $0 \vdash 100$, $100 \vdash 200$, $200 \vdash 300$, $300 \vdash 700$ si specifichi la distribuzione di frequenza assoluta e relativa e le corrispondenti frequenze cumulate. Si costruiscano le rappresentazioni grafiche mediante l'istogramma e il boxplot.

A.16) Una variabile quantitativa continua Y, osservata in tre diverse popolazioni indicate con A, B, C, fornisce le seguenti rilevazioni

$$\begin{aligned}
Y_A &= (-1.65, -5.74, -0.84, -4.50, 12.90, 0.84, -4.73, 7.89, 10.72, 4.40) \\
Y_B &= (-5.34, -6.26, 3.80, -2.15, -1.72, -7.35, 6.84, 5.92, 1.72, 7.76) \\
Y_C &= (7.05, -0.43, -10.21, -5.44, -6.75, -12.33, 13.76, 21.51, 0.43, 6.14)
\end{aligned}$$

Si confrontino i tre gruppi di dati, dopo aver calcolato media, mediana e varianza. Si dia una valutazione sulla eventuale simmetria. Si proponga una rappresentazione grafica comparata utilizzando i boxplot.

A.17) L'osservazione congiunta della variabili X e Y determina la seguente tabella delle frequenze relative congiunte

	X		
Y	0	1	2
2	0.10	0.30	0.15
4	0.20	0.10	0.15

Si determinino le distribuzioni di frequenza relativa marginali per X e per Y e la distribuzione condizionata di X dato $Y = 2$. Inoltre, si calcolino le associate medie aritmetiche.

A.18) In un quartiere si sono rilevati, in $n = 225$ abitazioni, il numero di televisori e il numero di persone che vi abitano. Si sono ottenuti i seguenti risultati, riassunti nella seguente tabella di contingenza

	Persone			
Televisori	1	2	3	4 o più
1	53	31	21	4
2 o più	27	49	25	15

Si determinino le distribuzioni di frequenza assoluta marginali per le due variabili. Si ottengano le corrispondenti frequenze relative congiunte e marginali. Si calcoli l'indice di connessione χ^2 e la sua versione normalizzata e si dia una valutazione sul livello di dipendenza tra le due variabili.

A.19) Nella tabella di contingenza presentata di seguito si riassumono i dati relativi al disastro del Titanic, suddividendo i passeggeri in base alla loro Classe di imbarco e al loro *Status* (morto o sopravvissuto)

	Classe			
Status	prima	seconda	terza	equipaggio
morto	122	167	528	673
sopravvissuto	203	118	178	212

Si determinino le distribuzioni di frequenza assoluta marginali per le due variabili. Si ottengano le corrispondenti frequenze relative congiunte e marginali e si definiscano le distribuzioni di frequenza relativa condizionata della variabile *Status*, rispetto ai valori della variabile Classe. Si calcoli l'indice di connessione χ^2 e la sua versione normalizzata e si dia una valutazione sul livello di dipendenza tra le due variabili.

A.20) Nella tabella di contingenza presentata di seguito si considerano i dati relativi al disastro del Titanic, suddividendo i passeggeri in base al Genere e al loro *Status* (morto o sopravvissuto)

	Genere	
Status	M	F
morto	1364	126
sopravvissuto	367	344

Si determinino le distribuzioni di frequenza assoluta marginali per le due variabili. Si ottengano le corrispondenti frequenze relative congiunte e marginali e si definiscano le distribuzioni di frequenza relativa condizionata della variabile *Status*, rispetto ai valori della variabile Genere. Si calcoli l'indice di connessione χ^2 e la sua versione normalizzata e si dia una valutazione sul livello di dipendenza tra le due variabili.

A.21) Viene misurata la lunghezza in *mm* di 15 uova di cuculo trovate in nidi di scricciolo e di 15 uova di cuculo trovate in nidi di pettirosso. Si ha rispettivamente

$$19.85, 20.05, 20.25, 20.85, 20.85, 20.85, 21.05, 21.05, 21.05, 21.25, 21.45, 22.05, 22.05, 22.05, 22.25$$

$$21.05, 21.85, 22.05, 22.05, 22.05, 22.25, 22.45, 22.45, 22.65, 23.05, 23.05, 23.05, 23.05, 23.05, 23.85$$

Indicata con Y la lunghezza e con X la tipologia di nido (scricciolo o pettirosso), si valuti la dipendenza in media tra Y e X e si rappresentino con i boxplot le distribuzioni di frequenza relativa condizionata di Y rispetto a X.

A.22) In Florida, si è condotto un esperimento per valutare se il trattamento delle nuvole con grandi quantità di ioduro di argento porta ad un incremento delle precipitazioni piovose. Si hanno i dati, in piedi per acri, sulle precipitazioni provenienti da 26 nuvole trattate e 26 non trattate

	Controllo	Trattamento		Controllo	Trattamento
1	1202.6	2745.6	14	41.1	200.7
2	830.1	1697.8	15	36.6	198.6
3	372.4	1656.0	16	29.0	129.6
4	345.5	978.0	17	28.6	119.0
5	321.2	703.4	18	26.3	118.3
6	244.3	498.1	19	26.1	115.3
7	163.0	430.0	20	24.4	92.4
8	147.8	334.1	21	21.7	40.6
9	95.0	302.8	22	17.3	32.7
10	87.0	274.7	23	11.5	31.4
11	81.2	274.7	24	4.9	17.5
12	68.5	255.0	25	4.9	7.7
13	47.3	242.5	26	1.0	4.1

Indicata con Y la quantità di precipitazioni e con X la presenza o meno del trattamento, si valuti la dipendenza in media tra Y e X e si rappresentino con i boxplot le distribuzioni di frequenza relativa condizionata di Y rispetto a X.

A.23) Con riferimento alla rilevazione di due variabili quantitative discrete X e Y, su $n = 100$ unità statistiche, si dispone delle seguente tabella con le frequenze assolute congiunte

	Y	
X	1	2
1	25	0
2	24	12
3	12	18
4	3	6

Si determinino le distribuzioni di frequenza assoluta marginali per le due variabili. Si ottengano le corrispondenti frequenze relative congiunte e marginali. Si specifichino le medie condizionate di Y dati i possibili valori per X. Infine, si calcolino le medie e le varianze marginali ed inoltre la covarianza e il coefficiente di correlazione lineare.

A.24) Si riportano i dati sulla spesa settimanale media per famiglia in sterline per gli alcolici (variabile esplicativa) e per i tabacchi (variabile risposta). Le osservazioni sono riferite all'anno 1981 e riguardano $n = 10$ regioni della Gran Bretagna

Regione	Alcolici	Tabacchi
1	6.47	4.03
2	6.13	3.76
3	6.19	3.77
4	4.89	3.34
5	5.63	3.47
6	4.52	2.92
7	5.89	3.20
8	4.79	2.71
9	5.27	3.53
10	6.08	4.51

Si determini la retta di regressione stimata e si valuti, calcolando l'indice di determinazione R^2, la bontà del modello. Si verifichi che $R^2 = \rho^2_{XY}$.

A.25) In $n = 10$ paesi dell'Unione Europea si è osservato il numero di veicoli circolanti pro-capite, variabile risposta Y, e il prezzo in euro di un litro di benzina, variabile esplicativa X. Si conoscono i seguenti valori di sintesi: $\sum_{i=1}^n x_i = 8.79$, $\sum_{i=1}^n y_i = 8.63$, $\sum_{i=1}^n x_i^2 = 7.738$, $\sum_{i=1}^n y_i^2 = 7.695$, $\rho_{XY} = -0.976$. Si determini la retta di regressione stimata e si valuti, calcolando l'indice di determinazione R^2, la bontà del modello.

A.26) Si sono osservati i seguenti $n = 14$ valori e variabili esplicativa X e risposta Y

X	2	2	2	2	4	4	5
Y	0.457	0.712	1.429	0.363	2.5	1.58	4.217
X	5	5	6	6	6	6	6
Y	2.521	3.022	3.235	3.666	2.739	3.644	3.93

Si determini la retta di regressione stimata e si valuti, calcolando l'indice di determinazione R^2, la bontà del modello.

A.27) Si sono osservati i seguenti $n = 15$ valori delle variabili esplicativa X e risposta Y

X	1.44	4.36	5.47	3.39	3.92	2.12	0.92	5.63
Y	10	798	816	42	155	51	3	832
X	5.76	2.72	2.71	3.62	4.05	1.63	0.24	
Y	1275	18	46	81	186	10	2	

Si determini la retta di regressione stimata e si valuti, calcolando l'indice di determinazione R^2, la bontà del modello. Si applichi di nuovo il modello di regressione dopo aver trasformato i valori della variabile risposta con il logaritmo naturale. Si dica quale dei due modelli di regressione risulta migliore.

Alcuni degli esercizi proposti sono tratti dai seguenti testi.
Navidi, W. (2006). *Probabilità e statistica per l'ingegneria e le scienze.* McGraw-Hill.
Pauli, F., Torelli, N. e Trevisani, M. (2008). *Statistica: esercizi ed esempi.* Pearson.
Triola, M.M. e Triola, F.T. (2009). *Statistica per le discipline biosanitarie.* Pearson.

Soluzioni

Si forniscono le soluzioni di alcuni quesiti presenti negli esercizi proposti

A.1) $y_{0.5} = 146$, $\mu = 146.8$, $y_{0.25} = 138$, $y_{0.75} = 155$, $R = 57$, $SI = 17$, $\sigma^2 = 170.32$, $CV = 0.089$.
A.2) Per le autovetture FIAT: $y_{mo} = 1$, $y_{0.5} = 1$, $\mu = 1.5$, $\sigma^2 = 1.4$, $CV = 0.789$. Per le autovetture OPEL: $y_{mo} = 0$, $y_{0.5} = 0$, $\mu = 0.58$, $\sigma^2 = 0.54$, $CV = 1.27$.
A.3) $y_{mo} = 75$, $y_{0.5} = 75$, $\mu = 75.25$, $y_{0.25} = 67.5$, $y_{0.75} = 82$, $R = 44$, $SI = 14.5$, $\sigma^2 = 107.63$, $CV = 0.138$.
A.4) Classe mediana $= 1200 \dashv 1300$, $\mu = 1321$, $\sigma^2 = 47159$.
A.5) Per la provincia di residenza, $y_{mo} =$ AL. Per il corso di laurea, $y_{mo} =$ C. Per il grado di soddisfazione, $y_{mo} = 3$, $y_{0.5} = 3$.
A.6) Classe mediana $= 65 \dashv 68$, $\mu = 66.93$, $\sigma^2 = 7.79$.
A.7) Per Ingegneria, classe mediana $= 27 \vdash 29$, $\mu = 28.89$, $\sigma^2 = 9.19$. Per Lettere, classe mediana $= 27 \vdash 29$, $\mu = 28.28$, $\sigma^2 = 11.93$.
A.8) Classe mediana $= 130 \dashv 160$, $\mu = 145.6$, $\sigma^2 = 1097.64$. Considerando i dati grezzi, $\mu = 147.1$, $y_{0.25} = 125$, $y_{0.5} = 148.5$, $y_{0.75} = 169$, $\min = 82$, $\max = 213$.
A.9) $y_{mo} = 1$, $y_{0.5} = 1$, $\mu = 1.65$, $\sigma^2 = 1.19$.
A.10) Classe mediana $= 75 \vdash 100$, $\mu = 88.97$, $\sigma^2 = 3157.41$.
A.11) $y_{mo} = 1$, $y_{0.5} = 2$, $\mu = 2.7$, $\sigma^2 = 4.04$
A.15) $\mu = 202.38$, $\sigma^2 = 14465, 76$, $y_{0.25} = 125$, $y_{0.5} = 170$, $y_{0.75} = 240$.
A.16) $\mu_A = 1.93$, $\mu_B = 0.32$, $\mu_C = 1.37$, $\sigma^2_A = 40.54$, $\sigma^2_B = 28.77$, $\sigma^2_C = 105.24$, $y^A_{0.25} = -4.61$, $y^A_{0.5} = 0$, $y^A_{0.75} = 9.30$, $y^B_{0.25} = -5.80$, $y^B_{0.5} = 0$, $y^B_{0.75} = 6.38$, $y^C_{0.25} = -8.48$, $y^C_{0.5} = 0$, $y^C_{0.75} = 10.40$.
A.17) X: 0.30, 0.40, 0.30; Y: 0.55, 0.45; $X \mid Y = 2$: 0.18, 0.55, 0.27; $\mu_X = 1$, $\mu_Y = 2.9$, $\mu_{X|Y=2} = 1.09$.
A.18) $\chi^2 = 19.02$, χ^2 normalizzato: 0.08.
A.19) $\chi^2 = 190.4$, χ^2 normalizzato: 0.09.
A.20) $\chi^2 = 456.97$, χ^2 normalizzato: 0.21.
A.21) Per $Y \mid X =$ scricciolo: $\mu = 21.13$, $y_{0.25} = 20.85$, $y_{0.5} = 21.05$, $y_{0.75} = 22.05$; per $Y \mid X =$ pettirosso: $\mu = 22.58$, $y_{0.25} = 22.05$, $y_{0.5} = 22.45$, $y_{0.75} = 23.05$.
A.22) Per $Y \mid X =$ Controllo: $\mu = 164.59$, $y_{0.25} = 23.05$, $y_{0.5} = 44.20$, $y_{0.75} = 203.65$; per $Y \mid X =$ Trattamento: $\mu = 442.33$, $y_{0.25} = 66.50$, $y_{0.5} = 221.60$, $y_{0.75} = 464.05$.
A.23) $\mu_{Y|X=1} = 1$, $\mu_{Y|X=2} = 1.33$, $\mu_{Y|X=3} = 1.6$, $\mu_{Y|X=4} = 1.67$, $\mu_X = 2.23$, $\mu_Y = 1.36$, $\sigma^2_X = 0.86$, $\sigma^2_Y = 0.23$, $Cov(X, Y) = 0.22$, $\rho_{XY} = 0.49$.
A.24) $\hat{a} = 0.11$, $\hat{b} = 0.61$, $R^2 = 0.614$, $\rho_{XY} = 0.784$.
A.25) $\hat{a} = 18.41$, $\hat{b} = -4.41$, $R^2 = 0.953$.
A.26) $\hat{a} = -0.628$, $\hat{b} = 0.7$, $R^2 = 0.819$.
A.27) Dati X e Y: $\hat{a} = -346.39$, $\hat{b} = 198.44$, $R^2 = 0.669$. Dati X e $\log(Y)$: $\hat{a} = 0.442$, $\hat{b} = 1.169$, $R^2 = 0.938$.

Suggerimento: provare a risolvere gli esercizi utilizzando anche le funzioni di R.

B) Calcolo delle probabilità

B.1) Si considerino dieci urne, identiche in apparenza, di cui nove contengono due palline bianche e due nere ciascuna e una contiene cinque palline bianche e una nera. Scelta a caso un'urna, si estrae una pallina che risulta essere bianca. Quale è la probabilità che la pallina sia stata estratta dall'urna contenente cinque palline bianche?

B.2) Si supponga di sapere che una porzione pari a 0.001 di una certa popolazione è affetto da una determinata malattia. Per rilevare tale patologia si utilizza un test diagnostico con le seguenti caratteristiche: se una persona è ammalata, il test risulta positivo con probabilità 0.999, mentre, se è sana il test risulta positivo con probabilità 0.002. Scelta a caso una persona, se il test risulta positivo, quale è la probabilità che la persona sia veramente affetta dalla malattia?

B.3) Si consideri un'urna contenente 10 palline, delle quali 5 sono nere. Si lanci un dado regolare e quindi si estragga dall'urna un campione, senza reinserimento, di ampiezza pari al numero che risulta dal lancio del dado. Determinare la probabilità che tutte le palline del campione siano nere.

B.4) Nel gioco del Superenalotto vengono estratti, senza reinserimento, da un'urna contenente i numeri da 1 a 90 sei numeri e un settimo numero detto numero jolly. Si totalizza 6, se si indovinano i sei numeri estratti; si totalizza $5+1$, se si indovinano 5 dei 6 numeri estratti e il numero jolly. Quale è la probabilità di totalizzare 6 scommettendo su sei numeri? Quale è la probabilità di totalizzare 6 o $5 + 1$ scommettendo su sei numeri? Quale è la probabilità di totalizzare 6 scommettendo su dieci numeri?

B.5) Una compagnia di assicurazione suddivide gli assicurati in persone propense e non propense agli incidenti. È noto che le persone propense agli incidenti hanno una probabilità pari a 0.2 di riportare incidenti in un anno, mentre questa probabilità scende a 0.05 per le altre persone. Se il 30% della popolazione è propensa agli incidenti, quale è la probabilità che un nuovo assicurato abbia un incidente nel primo anno? Se il nuovo assicurato ha effettivamente avuto un incidente nel primo anno, quale è la probabilità che sia propenso agli incidenti?

B.6) Un'industria alimentare si è dotata di un sistema automatico per il controllo della qualità. Tale sistema scarta un elemento non conforme con probabilità 0.9 e scarta erroneamente un elemento conforme con probabilità 0.01. Dall'analisi dei dati storici è noto che la proporzione di elementi non conformi prodotti dall'azienda è pari a 0.001. Se un elemento viene scartato dal sistema, quale è la probabilità che sia effettivamente non conforme?

B.7) Un gioco a premi televisivo è costituito da domande a risposta multipla. Con riferimento ad una singola domanda, sono previste 6 possibili risposte. Un concorrente decide di rispondere a caso se non conosce la risposta esatta. Sia 0.7 la probabilità che il concorrente conosca la risposta esatta alla singola domanda. Quale è la probabilità che il concorrente risponda esattamente ad una singola domanda? Se risponde correttamente alla domanda, quale è la probabilità che sia stato a conoscenza della risposta? Nell'ipotesi che il concorrente non modifichi la propria strategia di gioco, quale è la probabilità che su due domande risponda esattamente ad almeno una di esse?

B.8) Si considerano due urne. La prima contiene 3 palline nere e 7 bianche, mentre la seconda contiene 3 palline nere e 3 bianche. Si estrae una pallina dalla prima urna e, senza vederla, la si inserisce nella seconda. Successivamente si estrae dalla seconda urna una pallina. Quale è la probabilità che la pallina estratta dalla seconda urna sia bianca? Se la pallina estratta dalla seconda urna risulta essere bianca, quale è la probabilità che anche la pallina estratta dalla prima urna sia stata bianca?

B.9) Un sistema telegrafico trasmette linee e punti. È noto che 2/5 dei punti e 1/3 delle linee viene deformato durante la trasmissione. Inoltre, la probabilità che il sistema telegrafico trasmetta un punto è 5/8, mentre la probabilità che trasmetta una linea è 3/8. Si determini la probabilità che il segnale ricevuto sia uguale a quello trasmesso se il segnale ricevuto è un punto e se il segnale ricevuto è una linea.

B.10) Un'urna contiene dieci palline, tre nere e sette bianche. Ogni volta che si estrae una pallina dall'urna, viene reinserita con un'altra dello stesso colore. Quale è la probabilità di osservare pallina nera in tre estrazioni successive? Se alla seconda estrazione si è osservato pallina nera, quale è la probabilità che anche alla prima estrazione sia uscita pallina nera?

B.11) Una popolazione presenta il 32% di fumatori. È noto che il 25% dei fumatori e il 5% dei non fumatori è affetto da una patologia respiratoria cronica. Quale è la probabilità che un individuo scelto a caso da questa popolazione sia affetto dalla malattia? Quale è la probabilità che una persona scelta a caso da questa popolazione, e risultata affetta dalla malattia respiratoria, sia un fumatore?

B.12) Un fornitore di componenti elettronici raccoglie la sua produzione in tre classi, A, B, C, che contengono rispettivamente il 3%, il 9% e il 13% di pezzi difettosi. Le tre classi, inoltre, coprono rispettivamente il 15%, il 35% e il 50% della sua produzione. Avendo scelto un lotto di componenti, senza sapere a quale classe appartenga, e avendo riscontrato che un pezzo estratto a caso dal lotto è risultato difettoso, quale è la probabilità che il lotto in esame sia di classe C?

B.13) Si considerino 5 urne, identiche in apparenza, contenenti 4 palline nere e un numero variabile di palline bianche. In particolare, l'i-esima urna, $i = 1, \ldots, 5$, contiene $i + 2$ palline bianche. Si sceglie a caso un'urna e si estrae una pallina. Si calcoli la probabilità che la pallina estratta sia bianca. Supponendo che l'estrazione abbia dato come risultato una pallina bianca, si individui da quale urna è più plausibile che sia stata effettuata l'estrazione.

B.14) Uno studio sull'efficacia di un test diagnostico per l'anemia da carenza di ferro ha dato i seguenti risultati

	Anemia		
Test	Si	No	Totale
positivo	187	5	192
negativo	7	340	347
Totale	194	345	539

Sulla base di tali dati si stimi la probabilità che il test sia positivo se l'individuo è ammalato e la probabilità che il test sia negativo se l'individuo è sano. Se la probabilità che un individuo, scelto a caso dalla popolazione, sia affetto dalla patologia è 0.005, si calcoli la probabilità che l'individuo sia ammalato se il test è risultato positivo.

B.15) Una donna, sofferente di diabete mellito, insulino dipendente, alla 17ª settimana di gravidanza, è risultata positiva al test per l'alfa-proteina nel siero materno. La probabilità che il test sia positivo, se il feto risulta ammalato, e la probabilità che il test sia negativo, se il feto risulta sano, corrispondono a 0.34 e 0.86, rispettivamente. Se dalla letteratura medica si deduce che, in generale, c'è una probabilità pari a 0.20 che il feto sia ammalato, si calcoli la probabilità che ci sia la patologia se il test è risultato positivo.

B.16) Si consideri la funzione

$$F(x) = \begin{cases} 0 & \text{se } x < -1/2 \\ (1/2) + 2x + 2x^2 & \text{se } -1/2 \leq x < 0 \\ (1/2) + 2x - 2x^2 & \text{se } 0 \leq x < 1/2 \\ 1 & \text{se } x \geq 1/2. \end{cases}$$

Si verifichi che $F(x)$ è la funzione di ripartizione di una variabile casuale X continua e si ottenga la funzione di densità di probabilità. Si calcoli $P(|X| \leq 1/4)$, $P(X = 0)$ e $P(|X| \leq 1/4 | X \geq 0)$. Si determini il valore atteso, la moda, la mediana e la varianza di X.

B.17) Una rete è costituita da 15 PC per altrettanti operatori e da un server che permette la connessione di al più 10 PC. In un dato istante, ogni operatore richiede la connessione al server con probabilità $p = 0.5$. Ogni utente opera in modo indipendente. Quale è la probabilità che, ad un dato istante, la rete sia satura? Quale è il numero medio di operatori che, in un dato istante, si connettono al server?

B.18) Si consideri la funzione

$$F(x) = \begin{cases} a\exp(bx) & \text{se } x < 0 \\ 1 & \text{se } x \geq 0, \end{cases}$$

con $a \in [0, 1]$ e $b > 0$ opportuni numeri reali. Si dica se $F(x)$ è una funzione di ripartizione. Inoltre, si determini per quali valori di a e b $F(x)$ è una funzione di ripartizione di una variabile casuale continua. In questo caso, si calcoli la funzione di densità.

B.19) Sia X una variabile casuale con supporto $S_X = [2, +\infty)$ e funzione di densità $f_X(x) = kx^{-2}$, se $x \in S_X$, e nulla altrove, con k l'opportuna costante di normalizzazione. Si determini la costante k. Si calcoli il valore atteso, la moda, la mediana e la varianza di X.

B.20) Sia X una variabile casuale con supporto $S_X = [-1, +\infty)$ e funzione di densità $f_X(x) = ke^{-x}$, se $x \in S_X$, e nulla altrove, con k l'opportuna costante di normalizzazione. Si determini la costante k. Si calcoli il valore atteso, la moda e la mediana di X.

B.21) Su mille automobili vendute di un certo modello, ogni mese 10 richiedono un intervento di assistenza. Un concessionario ha venduto 100 automobili di questo tipo. Si proponga un modello probabilistico ragionevole per la variabile casuale X che conta il numero di interventi di assistenza (alle 100 automobili vendute) che tale concessionario effettuerà nel prossimo mese. Si calcolino $E(X)$, $V(X)$ e $P(X = 0)$.

B.22) Sia X la variabile casuale che conta il numero di volte con cui esce un numero maggiore o uguale a 5, lanciando tre volte un dado. Nell'ipotesi che il dado sia regolare, si determini il supporto, la funzione di probabilità e la funzione di ripartizione di X. Inoltre, si calcoli il valore atteso, la moda, la mediana e la varianza di X.

B.23) Sia X la variabile casuale che conta il numero di lanci di un dado affinché esca per la prima volta un numero maggiore o uguale a 5. Nell'ipotesi che il dado sia regolare, si proponga un modello probabilistico ragionevole per la variabile casuale X. Inoltre, si calcoli il valore atteso e la varianza di X.

B.24) Un test di ingresso è costituito da 4 domande a risposta multipla, ognuna delle quali ha 5 risposte possibili, con una sola corretta. Lo studente impreparato risponde casualmente ad ognuna delle domande. Si proponga un modello probabilistico ragionevole per la variabile casuale X che conta il numero di risposte positive alle 4 domande da parte dello studente. Si calcolino $E(X)$, $V(X)$, $P(X = 3)$ e $P(X = 0)$.

B.25) In una piccola concessionaria di automobili si presenta mediamente un cliente ogni 4 ore. Si proponga un modello probabilistico ragionevole per la variabile casuale X che conta il numero di clienti che si presentano in un ora. Si calcoli la probabilità che in un'ora entri esattamente una persona e che in un'ora entri almeno una persona.

B.26) Il tempo che mediamente trascorre tra l'apertura di uno sportello pubblico e l'arrivo del primo utente è di 4 minuti. Si proponga un modello probabilistico ragionevole per la variabile casuale X che descrive il tempo tra l'apertura dello sportello e l'arrivo del primo utente. Si determini la probabilità che il tempo d'attesa del primo utente superi i 5 minuti. Infine, si calcolino i quartili di X.

B.27) Un'azienda fabbrica termometri che si suppongono in grado di fornire la lettura di 0°C al punto di congelamento dell'acqua. In realtà, la lettura che forniscono è descritta da una variabile casuale $X \sim N(0.05, 1.1)$. Scelto a caso un termometro, determinare la probabilità che, al punto di congelamento dell'acqua, la lettura sia meno di 1.58°C, sia almeno -1.23°C e sia compresa tra -2.00°C e 1.50°C. Infine, si determinino i valori di temperatura che delimitano il 2.5% inferiore e il 2.5% superiore della distribuzione di probabilità si X.

B.28) La durata di un circuito integrato è descritta da una variabile casuale con distribuzione esponenziale con media pari 2 anni. Si determini la probabilità che il circuito duri più di 3 anni. Inoltre, nell'ipotesi che il circuito dopo 4 anni sia ancora funzionante, si calcolo la probabilità che funzioni per almeno altri 3 anni.

B.29) La variabile casuale X, che descrive il numero di difetti in un filo di rame di lunghezza 3 cm, ha la seguente distribuzione di probabilità

x	0	1	2	3
$P(X = x)$	0.48	0.39	0.12	0.01

Si considerano 100 fili di rame. Utilizzando una opportuna approssimazione, si determini la probabilità che il numero medio di difetti per filo, considerando i 100 fili in esame, sia minore di 0.5 e la probabilità che la somma totale dei difetti riscontrati nei 100 fili sia minore di 25.

B.30) Il numero medio di connessioni ad un sito web segue una distribuzione di poisson con una media di 27 connessioni all'ora. Si determini la probabilità che avvengano più di 90 connessioni in tre ore.

B.31) Nella fase di progettazione dei sedili di un aereo commerciale si vuole considerare una larghezza tale da soddisfare la quasi totalità dei potenziali utilizzatori maschi. Tenendo conto del fatto che la larghezza in cm dei fianchi degli uomini è descritta da una variabile casuale $X \sim N(36.5, 6.25)$, si determini l'ampiezza dei sedili che li rende accessibili al 95% degli uomini.

B.32) È noto che il 5% dei semi di frumento di una determinata partita non germoglierà. I semi vengono venduti in confezioni da 200 semi, dei quali viene garantita la germinazione per almeno il 90%. Utilizzando una opportuna approssimazione, si determini la probabilità che una confezione non sia conforme a tale garanzia.

B.33) Sia $\{X_n\}_{n\geq1}$ una successione di variabili casuali con distribuzione di probabilità poisson di parametro $n\lambda$, con $\lambda > 0$. Si studi la convergenza in probabilità della successione $\{Y_n\}_{n\geq1}$, con $Y_n = X_n/n$. Avendo standardizzato opportunamente le variabili casuali della successione $\{Y_n\}_{n\geq1}$, si studi la convergenza in distribuzione. Si calcoli in modo approssimato $P(X_n > n\lambda)$, con n sufficientemente elevato.

B.34) Un dado equilibrato viene lanciato 900 volte. Sia X la variabile casuale che conta il numero di volte in cui compare il numero 6. Determinare $E(X)$, $V(X)$ e, con una opportuna approssimazione, $P(X \geq 180)$. Per un difetto di fabbricazione, una partita di dadi contiene alcuni non equilibrati che, in particolare, associano al numero 6 una probabilità pari a 2/9. Si decide di considerare truccato un dado se, lanciato 900 volte, presenta il 6 in almeno 180 lanci. Quale è la probabilità che, adottando questa procedura, si riesca ad individuare un dado non equilibrato?

B.35) Siano $X_1, \ldots, X_n$, $n \geq 1$, variabili casuali indipendenti con distribuzione di probabilità esponenziale di parametro $\lambda > 0$. Si consideri la variabile casuale S_n che conta il numero delle variabili $X_1, \ldots, X_n$ la cui realizzazione sarà più grande del loro valore atteso. Si determini la distribuzione di probabilità di S_n. Si studi la convergenza in probabilità della successione $\{Y_n\}_{n\geq1}$, dove $Y_n = S_n/n$.

Alcuni degli esercizi proposti sono tratti dai seguenti testi.

Navidi, W. (2006). *Probabilità e statistica per l'ingegneria e le scienze.* McGraw-Hill.
Pauli, F., Torelli, N. e Trevisani, M. (2008). *Statistica: esercizi ed esempi.* Pearson.
Triola, M.M. e Triola, F.T. (2009). *Statistica per le discipline biosanitarie.* Pearson.

Soluzioni

Si forniscono le soluzioni di alcuni quesiti presenti negli esercizi proposti

B.1) 0.156.
B.2) 0.33.
B.3) 0.138.
B.4) $1.606 \cdot 10^{-9}$, $1.124 \cdot 10^{-8}$, $3.373 \cdot 10^{-7}$.
B.5) 0.095, 0.632.
B.6) 0.083.
B.7) 0.75, 0.933, 0.937.
B.8) 0.529, 0.757.
B.9) 0.75, 0.5.
B.10) 0.045, 0.364.
B.11) 0.114, 0.702.
B.12) 0.64.
B.13) 0.544, urna 5.
B.14) 0.964, 0.985, 0.244.
B.15) 0.378.
B.16) 0.75, 0, 0.75, 0, 0, 0, 0.042.
B.17) 0.15, 7.5.
B.18) $a = 1$, $b > 0$.
B.19) 2, $E(X)$ e $V(X)$ non esistono, 2, 4.
B.20) e^{-1}, 0, -1, $\log 2 - 1$.
B.21) 1, 0.99 0.366.
B.22) 1, 1, 1, 0.667.
B.23) 3, 6.
B.24) 0.8, 0.64, 0.026, 0.41.
B.25) 0.195, 0.221.
B.26) 0.286, 1.15, 2.77, 5.54.
B.27) 0.927, 0.889, 0.891, -2.01, 2.11.
B.28) 0.223, 0.223.
B.29) 0.014, $7.57 \cdot 10^{-9}$.
B.30) 0.159.
B.31) 40.625.
B.32) 0.006.
B.33) λ, 0.5.
B.34) 150, 125, 0.0037, 0.95.
B.35) e^{-1}.

Suggerimento: provare a risolvere gli esercizi utilizzando anche le funzioni di R.

C) Inferenza statistica

C.1) Si consideri un campione casuale semplice costituito da variabili casuali $X_1, \ldots, X_n$, $n \geq 1$, indipendenti con distribuzione di probabilità geometrica di parametro $p \in (0, 1)$ ignoto. Si determini un opportuno stimatore per p e si dica se è consistente. Si esprima un parere sulla eventuale non distorsione di tale stimatore.

C.2) Si consideri un campione casuale semplice costituito da variabili casuali $X_1, \ldots, X_n$, $n \geq 1$, indipendenti con funzione di densità $f(x; \theta) = 3x^2/\theta^3$, se $x \in [0, \theta]$, e nulla altrove, dove il parametro $\theta > 0$ è ignoto. Si individui un opportuno stimatore per θ e si dica se è non distorto e consistente. Inoltre, si determini lo *standard error* di tale stimatore.

C.3) Si consideri un campione casuale semplice costituito da variabili casuali $X_1, \ldots, X_n$, $n \geq 1$, indipendenti con distribuzione di probabilità binomiale di parametri $m = 10$ e $p \in (0, 1)$ ignoto. Si individui un opportuno stimatore per p e si dica se è non distorto e consistente. Inoltre, si determini lo *standard error* di tale stimatore. Nell'ipotesi che il campione osservato sia $(2, 8, 1, 4, 5, 5, 3, 6, 3, 3)$, si calcoli la stima per p e lo *standard error* stimato.

C.4) In una sperimentazione in ambito agrario si vuole stimare la probabilità p di successo connessa all'utilizzazione di un nuovo trattamento antiparassitario. Quale deve essere la dimensione n del campione affinché l'errore di stima sia, in valore assoluto, inferiore a 0.1 con probabilità 0.90? Dal momento che il risultato richiede la conoscenza di p, si sostituisca a $p(1-p)$ la sua limitazione superiore $1/4$, che porta ad una valutazione prudente per n.

C.5) Il raggio μ di un cerchio non è noto e viene misurato più volte con uno strumento che fornisce valori affetti da errore. Si ipotizza che le n osservazioni siano variabili casuali indipendenti con distribuzione di probabilità normale di media $\mu \in \mathbf{R}$ e varianza $\sigma^2 > 0$, entrambe ignote. Si determini uno stimatore per l'area del cerchio e si verifichi l'eventuale non distorsione. Se lo stimatore risulta distorto, si proponga una opportuna versione modificata che definisca uno stimatore non distorto.

C.6) I dati $(0, 0, 3, 0, 1, 0, 0, 2, 1, 0, 0, 2)$, relativi ad osservazioni ripetute del numero di visite ad un sito web in una fissata unità di tempo, sono analizzati come realizzazione di un campione casuale semplice $X_1, \ldots, X_{12}$, costituito da variabili casuali indipendenti con distribuzione di probabilità poisson di parametro $\lambda > 0$ ignoto. Si individui uno stimatore ragionevole per λ e si verifichi se è non distorto. Si calcoli la stima di λ e l'associato *standard error* stimato.

C.7) Si consideri un campione casuale semplice costituito da variabili casuali $X_1, \ldots, X_n$, $n \geq 1$, copie indipendenti di $X \sim N(\mu, \sigma^2)$, con μ e σ^2 ignoti. Si introduca un opportuno stimatore per $\theta = E(X^2)$ e si dica se è non distorto e consistente. Se lo stimatore risulta distorto, si proponga una opportuna versione modificata che definisca uno stimatore non distorto.

C.8) Si consideri un campione casuale semplice costituito da variabili casuali X_1, X_2, X_3, con valore atteso e varianza pari a λ, non noto. Si definiscono i seguenti stimatori per λ: $\hat{\lambda}_1 =$

$(2X_1 + X_2 + 2X_3)/5$ e $\hat{\lambda}_2 = (X_1 + 2X_2 + X_3)/4$. Si verifichi se tali stimatori sono non distorti e si dica quale dei due è preferibile.

C.9) Nell'ambito di una procedura per il controllo di qualità, si considera un campione casuale semplice di 1000 pezzi prodotti in un dato giorno e si riscontra che 30 di essi sono difettosi. Si definisca un modello statistico parametrico utile per descrivere l'esperimento in esame. Inoltre, si introduca un opportuno stimatore per la proporzione p di pezzi difettosi, si determini l'associato *standard error* e si calcolino i corrispondenti valori di stima. Inoltre, si definisca un intervallo di confidenza per p, con livello di confidenza approssimato $1 - \alpha = 0.95$. Se si ripete l'esperimento per 260 giorni e si costruiscono 260 intervalli di confidenza come sopra, quanti di questi intervalli ci si attende includano il vero valore di p?

C.10) Si consideri un campione casuale semplice costituito da variabili casuali $X_1, \ldots, X_n$, $n \geq 1$, indipendenti con distribuzione di probabilità esponenziale di parametro $\lambda > 0$ ignoto. Si introduca un opportuno stimatore per $\mu = 1/\lambda$ e si determini l'associato *standard error*. Nell'ipotesi che, con riferimento ai risultati dell'indagine campionaria, si abbia $n = 100$ e $\sum_{i=1}^{100} x_i = 187.44$, si fornisca una stima per μ, si calcoli l'associato *standard error* stimato e si ottenga un intervallo di confidenza per μ di livello approssimato $1 - \alpha = 0.95$.

C.11) Si consideri un campione casuale semplice costituito da variabili casuali $X_1, \ldots, X_n$, $n \geq 1$, indipendenti con distribuzione di probabilità normale di media $\mu \in \mathbf{R}$ e varianza $\sigma^2 > 0$. Si calcolino gli intervalli di confidenza per μ di livello $1 - \alpha = 0.9$ e $1 - \alpha = 0.95$ (a) nell'ipotesi che, con $\sigma^2 = 25$ noto, $n = 30$ e $\sum_{i=1}^{30} x_i = 88.2$ e (b) nell'ipotesi che, con σ^2 ignoto, $n = 30$, $\sum_{i=1}^{30} x_i = 88.2$ e $\sum_{i=1}^{30}(x_i - 2.94)^2 = 148.5$.

C.12) Si consideri un campione casuale semplice costituito da variabili casuali $X_1, \ldots, X_{100}$, che corrispondono alla misurazione del tasso di colesterolo nel sangue di 100 individui, scelti a caso da una certa popolazione oggetto di studio. L'esperimento ha fornito i seguenti risultati in g/dl: $\sum_{i=1}^{100} x_i = 155$ e $\sum_{i=1}^{100} x_i^2 = 290.25$. Nell'ipotesi che il modello statistico di riferimento sia normale, si fornicano gli intervalli di confidenza di livello $1 - \alpha = 0.90$ per la media e la varianza del tasso di colesterolo nel sangue.

C.13) Un'industria alimentare che produce pasta dispone di un macchinario che fornisce pacchi dal peso netto dichiarato di 0.5 Kg. In realtà, il peso di una generica confezione di pasta è descritto da una variabile casuale con distribuzione di probabilità normale di media $\mu \in \mathbf{R}$ e varianza $\sigma^2 > 0$, entrambe ignote. Si estraggono casualmente 12 pacchi di pasta, il cui peso effettivo in Kg è $(0.498, 0.489, 0.503, 0.493, 0.491, 0.499, 0.512, 0.504, 0.483, 0.506, 0.510, 0.509)$. Sulla base di tale campione osservato, si determinino gli intervalli di confidenza per μ di livello $1 - \alpha = 0.95$ e $1 - \alpha = 0.99$. Inoltre, si calcolino gli intervalli di confidenza per σ^2 di livello $1 - \alpha = 0.9$ e $1 - \alpha = 0.99$.

C.14) Si consideri un campione casuale semplice costituito da variabili casuali $X_1, \ldots, X_n$, $n \geq 1$, indipendenti con distribuzione di probabilità poisson di parametro $\lambda > 0$ ignoto. Nell'ipotesi che n sia sufficientemente elevato, si determini un intervallo di confidenza approssimato per λ di livello

$1-\alpha=0.95$. Si calcoli tale intervallo di confidenza con riferimento ad un campione osservato di dimensione $n=90$ tale che $\sum_{i=1}^{90} x_i = 135$.

C.15) Si consideri un campione casuale semplice costituito da variabili casuali $X_1, \dots, X_n$, $n \geq 1$, indipendenti con distribuzione di probabilità binomiale di parametri $m=10$ e $p \in (0,1)$ ignoto. Nell'ipotesi che n sia sufficientemente elevato, si determini un intervallo di confidenza approssimato per p di livello $1-\alpha=0.95$. Si calcoli tale intervallo di confidenza con riferimento ad un campione osservato di dimensione $n=100$, tale che $\sum_{i=1}^{100} x_i = 389$.

C.16) È noto, dall'esperienza passata, che il numero di accessi in un'ora ad un sito per il commercio elettronico è descritto da una variabile casuale con distribuzione di probabilità poisson di parametro λ ignoto. Si consideri in campione casuale semplice $X_1, \dots, X_n$ che descrive il numero di accessi nelle ultime $n \geq 1$ ore; la variabile casuale X_i, $i=1,\dots,n$, corrisponde al numero di accessi nell'i-esima ora. Si introduca un opportuno stimatore per λ e si determini l'associato *standard error*. Se durante le ultime $n=96$ ore si sono verificati 383 accessi, si fornisca, sulla base dell'evidenza empirica, una stima per λ e si calcoli l'associato *standard error* stimato. Inoltre, si determini un intervallo di confidenza per λ di livello approssimato $1-\alpha=0.90$.

C.17) Si consideri un campione casuale semplice costituito da variabili casuali $X_1, \dots, X_n$, $n \geq 1$, indipendenti con distribuzione di probabilità normale con valor medio μ e varianza σ^2 ignoti. Si introduca un opportuno stimatore per σ^2 e si individuino le sue proprietà. Nell'ipotesi che, con riferimento ai risultati dell'indagine campionaria, si abbia $n=30$, $\sum_{i=1}^{30} x_i = 36.76$ e $\sum_{i=1}^{30} x_i^2 = 102.27$, si ottenga un intervallo di confidenza per σ^2 di livello $1-\alpha=0.95$.

C.18) Si consideri un campione casuale semplice costituito da variabili casuali $X_1, \dots, X_n$, $n \geq 1$, indipendenti con distribuzione di probabilità normale di media $\mu \in \mathbf{R}$ ignota e varianza $\sigma^2 = 4$ nota. Si calcolino le regioni di rifiuto unilaterale destra, unilaterale sinistra e bilaterale, di livello $\alpha=0.10$, $\alpha=0.05$ e $\alpha=0.01$, per l'ipotesi nulla $H_0: \mu=20$, nel caso in cui $n=10$. Nell'ipotesi che $\sum_{i=1}^{10} x_i = 185.8$, si tragga una conclusione sull'ipotesi nulla, con riferimento a ciascuno dei tre livelli di confidenza. Infine, si determini il livello di significatività osservato nei tre casi di ipotesi alternativa unilaterale destra, unilaterale sinistra e bilaterale.

C.19) Un'industria alimentare che produce birra dichiara che il contenuto nominale di ogni bottiglia è 330 ml. In realtà, il contenuto di una generica bottiglia è descritto da una variabile casuale con distribuzione di probabilità normale di media $\mu \in \mathbf{R}$ e varianza $\sigma^2 > 0$, entrambe ignote. Si scelgono casualmente 20 bottiglie, il cui contenuto totale osservato è $\sum_{i=1}^{20} x_i = 6560$, con $\sum_{i=1}^{20}(x_i - 328)^2 = 194.56$. Sulla base di tale campione osservato, si verifichi l'ipotesi nulla $H_0: \mu=330$, al livello $\alpha=0.05$, contro un'ipotesi alternativa unilaterale sinistra. Inoltre, si calcoli il livello di significatività osservato.

C.20) È noto, dall'esperienza passata, che il numero di interruzioni che ogni settimana avvengono in una determinata linea di produzione è descritto da una variabile casuale con distribuzione di probabilità poisson di parametro $\lambda=1$. Se durante le ultime 52 settimane si sono verificate 85 interruzioni, si può ragionevolmente affermare che l'entità del fenomeno è cresciuta in modo significativo? Per rispondere a tale domanda, si verifichi l'ipotesi nulla $H_0 : \mu = 1$, al livello $\alpha = 0.01$ approssimato, contro un'ipotesi alternativa unilaterale destra. Inoltre, si calcoli il livello di significatività osservato approssimato.

C.21) Si consideri un campione casuale semplice costituito da variabili casuali $X_1, \ldots, X_n$, $n \geq 1$, indipendenti con distribuzione di probabilità normale di media $\mu \in \mathbf{R}$ e varianza σ^2 ignota. Si calcolino le regioni di rifiuto unilaterale destra, unilaterale sinistra e bilaterale, di livello $\alpha = 0.05$ e $\alpha = 0.01$, per l'ipotesi nulla $H_0 : \sigma^2 = 12.5$, nel caso in cui $n = 12$. Nell'ipotesi che si sia osservato il campione $(1, 2, 4, 5, 3, 5, 6, 0, 4, 6, 5, 8)$, si tragga una conclusione sull'ipotesi nulla, con riferimento ai due livelli di confidenza. Infine, si determini il livello di significatività osservato nel caso di regioni di rifiuto unilaterale destra, unilaterale sinistra e bilaterale.

C.22) Da indagini precedentemente svolte è noto che il peso alla nascita in Kg dei neonati, in una determinata regione italiana, può essere ragionevolmente descritto da una variabile casuale normale con varianza 0.36. Si considera un campione di 20 neonati con madre fumatrice e, avendo rilevato il peso alla nascita $x_1, \ldots, x_{20}$, si ricava che $\bar{x}_{20} = 3.2$ e $(1/20)\sum_{i=1}^{20} x_i^2 = 10.69$. Si vuole indagare l'eventualità che l'abitudine al fumo porti ad un incremento nella variabilità del peso alla nascita. Per rispondere a tale interrogativo si verifichi l'ipotesi $H_0 : \sigma^2 = 0.36$, a fronte di un'alternativa unilaterale destra, considerando un livello di significatività $\alpha = 0.05$.

C.23) Si considera un lotto di lampadine a basso consumo e si vuole verificare se la proporzione p di lampadine difettose è 0.03, come affermato dal produttore. A tal fine si scelgono a caso 200 lampadine e 12 di esse risultano difettose. Si definisca un modello statistico parametrico utile per descrivere l'esperimento in esame, si introduca un opportuno stimatore per p, si determini l'associato *standard error* e si calcolino i corrispondenti valori di stima. Infine, con riferimento ai risultati dell'indagine campionaria, si verifichi l'ipotesi $H_0 : p = 0.03$ contro un'alternativa unilaterale destra, ad un livello di significatività (approssimato) $\alpha = 0.05$.

C.24) In un'indagine sull'altezza della popolazione adulta maschile si sono considerati due campioni casuali semplici, di dimensione 100 e 50 rispettivamente, da tratti dalla popolazione residente in due distinte regioni. Dai dati campionari si ottiene che l'altezza media in cm degli individui estratti dalla prima popolazione è 170.3 e la media delle altezze al quadrato corrisponde a 29120. Per quanto riguarda gli individui estratti dalla seconda popolazione, si ottengono i seguenti valori di sintesi: 166.8 e 27943. Assumendo che l'altezza segua, in entrambe le regioni, una distribuzione normale con la stessa varianza, si verifichi l'ipotesi che le altezze medie siano uguali, a fronte di un'alternativa bilaterale, considerando un livello di significatività $\alpha = 0.05$.

C.25) Si vogliono confrontare due metodi alternativi per determinare il contenuto di nichel nell'acciaio. Si considerano 12 campioni di acciaio provenienti dalla stessa partita; con riferimento a 5 di essi si utilizza la prima metodologia, mentre con riferimento agli altri 7 si applica la seconda metodologia. Nel primo caso si ottengono i seguenti valori osservati per la media e la varianza campionaria: 3.16 e 0.0018. Nel secondo caso, si ottengono i valori 3.24 e 0.0023. Assumendo che le misurazioni, con entrambi i metodi, seguano una distribuzione normale con la stessa varianza, si verifichi l'ipotesi che in media le due metodologie forniscano misurazioni uguali, a fronte di un'alternativa bilaterale, considerando un livello di significatività $\alpha = 0.01$.

C.26) Il Captopril è un farmaco per abbassare la pressione arteriosa. Per valutare la sua efficacia si considera un campione casuale semplice di 12 individui a cui viene misurata la pressione arteriosa prima e dopo l'assunzione del farmaco. I dati vengono riportati nella seguente tabella

Individuo	1	2	3	4	5	6	7	8	9	10	11	12
Prima	200	174	198	170	179	182	193	209	185	155	169	210
Dopo	191	170	177	167	159	151	176	183	159	145	146	177

Con una opportuna procedura di verifica di ipotesi, si valuti se, alla luce del campione osservato, il Captopril sembra efficace per ridurre la pressione arteriosa, considerando un livello di significatività $\alpha = 0.01$.

C.27) Per verificare l'efficacia di un vaccino in spray nasale per bambini, lo si somministra ad un campione di 1070 bambini, mentre ad un secondo campione di 532 bambini si somministra un placebo. Nel primo caso si riscontra che 14 bambini hanno contratto l'influenza, mentre nel secondo caso gli ammalati risultano 95. Con una opportuna procedura di verifica di ipotesi, si valuti se, alla luce dei campioni osservati, la proporzione di casi di influenza tra i bambini vaccinati risulti significativamente inferiore alla proporzione di casi di influenza tra i bambini che hanno ricevuto il placebo, considerando un livello di significatività $\alpha = 0.01$.

C.28) Per studiare l'eventuale effetto degli integratori di calcio sulla pressione sanguigna, si è considerato un campione di pazienti di dimensione 13 a cui si è somministrato un placebo e un un campione di pazienti di dimensione 15, trattato con gli integratori. I dati campionari sono, rispettivamente

Placebo	124.6	104.8	96.5	116.3	106.1	128.8	107.2	123.1	118.1	108.5	120.4
	122.5	113.6									
Calcio	129.1	123.4	102.7	118.1	114.7	120.9	104.4	116.3	109.6	127.7	108.0
	124.3	106.6	121.4	113.2							

Nell'ipotesi che il modello statistico parametrico sia Gaussiano, si verifichi ad un livello di significatività $\alpha = 0.05$, l'ipotesi che i due campioni provengano da popolazioni con la stessa varianza.

C.29) Per studiare il tasso di assorbimento di due pesticidi sulla pelle si sono considerati due campioni di dimensione 6 e 10 che fanno riferimento alla quantità di pesticida, in μg, assorbita dalla pelle, nel caso si consideri, rispettivamente, il primo e il secondo pesticida. Dai dati campionari si ricava che, nel primo caso la varianza è 2.3, mentre nel secondo caso è 0.6. Nell'ipotesi che il modello statistico parametrico sia Gaussiano, si verifichi ad un livello di significatività $\alpha = 0.01$, l'ipotesi che i due campioni provengano da popolazioni con la stessa varianza.

C.30) Gli scambi di dati tra due computer possono avvenire utilizzando due diversi protocolli di instradamento. Si verifica che, utilizzando il primo protocollo, su 200 messaggi inviati, 170 vengono ricevuti correttamente, mentre con il secondo protocollo, su 150 messaggi inviati, 123 vengono ricevuti correttamente. Si può ragionevolmente affermare, ad un livello di significatività (approssimato) $\alpha = 0.05$, che il primo protocollo ha un tasso di successo maggiore del secondo? Si calcoli, inoltre, il livello di significatività osservato approssimato.

C.31) Da un'indagine condotta su 100 possessori di carta di credito emerge che 57 affermano di sapere che con l'utilizzo della carta di credito si possono acquisire punti fedeltà. Dopo lo svolgimento di una campagna pubblicitaria volta a far conoscere la promozione basata sui punti fedeltà, si conduce una nuova inchiesta da cui risulta che su 200 possessori di carta di credito 135 affermano di conoscere la promozione. Si può ragionevolmente affermare, ad un livello di significatività (approssimato) $\alpha = 0.01$, che la campagna pubblicitaria è stata efficace? Si calcoli, inoltre, il livello di significatività osservato approssimato.

C.32) Per valutare l'efficacia di determinate scatole di imballaggio pesanti, vengono spediti 1200 prodotti con un imballaggio leggero e 1500 prodotti con un imballaggio pesante. Nel primo caso si verifica che 20 prodotti risultano danneggiati, mentre nel secondo caso i prodotti danneggiati sono 15. Si può ragionevolmente affermare, ad un livello di significatività (approssimato) $\alpha = 0.01$, che l'imballaggio pesante riduce la proporzione di prodotti danneggiati? Si calcoli, inoltre, il livello di significatività osservato approssimato.

C.33) In una industria meccanica ci sono quattro macchinari che producono perni cilindrici di acciaio. I perni, che devono rispettare determinate specifiche di diametro, vengono classificati in "Fini", "Ok", "Spessi". Si scelgono casualmente dei perni tra quelli prodotti dai quattro macchinari e vengono classificati nelle tre categorie, come sintetizzato nella seguente tabella,

	Fini	Ok	Spessi	Totale
Macchinario 1	10	102	8	120
Macchinario 2	34	161	5	200
Macchinario 3	12	79	9	100
Macchinario 4	10	60	10	80
Totale	66	402	32	500

Si valuti, utilizzando un opportuno test di ipotesi, se la qualità dei perni dipende dal macchinario che li ha prodotti, ad un livello di significatività (approssimato) $\alpha = 0.05$.

C.34) Un'industria meccanica produce perni cilindrici di acciaio, che devono rispettare determinate specifiche di diametro e di lunghezza. Con riferimento al diametro, vengono classificati in "Fini", "Ok", "Spessi", mentre con rifermento alla lunghezza, vengono classificati in "Corti", "Ok", "Lunghi". Si scelgono casualmente 1021 perni tra quelli prodotti e vengono classificati rispetto alle specifiche di diametro e di lunghezza, come sintetizzato nella seguente tabella,

	Fini	Ok	Spessi	Totale
Corti	13	117	4	134
Ok	62	664	80	806
Lunghi	5	68	8	81
Totale	80	849	92	1021

Si valuti, utilizzando un opportuno test di ipotesi, se c'è dipendenza tra lunghezza e diametro, ad un livello di significatività (approssimato) $\alpha = 0.01$.

C.35) Da un indagine sull'occupazione condotta su un campione di 232 individui si ottiene la seguente tabella di frequenze relative osservate

	Laureato	Non laureato	Totale
Occupato	0.24	0.51	0.75
Non stabile	0.055	0.095	0.15
Disoccupato	0.005	0.095	0.10
Totale	0.30	0.70	1

Si valuti, utilizzando un opportuno test di ipotesi, se c'è dipendenza tra occupazione e titolo di studio, ad un livello di significatività (approssimato) $\alpha = 0.05$.

C.36) Lanciando 100 volte un dado a sei facce si è ottenuto 20 volte la faccia 1, 17 volte la faccia 2, 13 volte la faccia 3, 13 volte la faccia 4, 23 volte la faccia 5 e 14 volte la faccia 6. Si dica se, sulla base di un opportuno test di ipotesi con livello di significatività (approssimato) $\alpha = 0.05$, il dado può ritenersi truccato.

C.37) Da un'indagine sui tempi di reazione in ms agli stimoli visivi X e uditivi Y, condotta su un campione di 10 soggetti, si ottengono i seguenti risultati

x	161	203	235	176	201	188	228	211	191	178
y	159	206	241	163	197	193	209	189	169	201

Si definisca uno stimatore per il coefficiente di correlazione lineare tra X e Y e si determini l'associato valore di stima. Inoltre, utilizzando un opportuno test di ipotesi, si verifichi l'ipotesi di assenza di correlazione a fronte di un'ipotesi alternativa unilaterale destra, ad un livello di significatività $\alpha = 0.01$.

C.38) Con riferimento a 10 campioni di rame si misurano la resistenza alla trazione X e la durezza Y. I dati vengono riportati nella seguente tabella

x	106.2	106.3	105.3	106.1	105.4	106.3	104.7	105.4	105.5	105.1
y	35.0	37.2	39.8	35.8	41.3	40.7	38.7	40.2	38.1	41.6

Si definisca uno stimatore per il coefficiente di correlazione lineare tra X e Y e si determini l'associato valore di stima. Inoltre, utilizzando un opportuno test di ipotesi, si verifichi l'ipotesi di assenza di correlazione a fronte di un'ipotesi alternativa bilaterale, ad un livello di significatività $\alpha = 0.05$.

C.39) Si vuole studiare l'effetto che un certo additivo ha sul tempo di asciugatura di un determinato capo d'abbigliamento. Si effettuano le seguenti $n = 10$ misurazioni sulla variabile esplicativa x, che descrive la concentrazione dell'additivo in %, e sulla variabile risposta Y, che corrisponde al tempo di asciugatura in ore

x	4.0	4.2	4.4	4.6	4.8	5.0	5.2	5.4	5.6	5.8
y	8.7	8.8	8.3	8.7	8.1	8.0	8.1	7.7	7.5	7.2

Nell'ipotesi che il modello di regressione lineare semplice sia appropriato, si determini la retta di regressione stimata. Con riferimento agli stimatori per l'intercetta e per il coefficiente angolare, si determinino gli standard error e si calcolino gli associati standard error stimati. Inoltre, si proponga una stima per la varianza di Y. Si verifichi l'ipotesi che il coefficiente angolare della retta di regressione sia nullo, a fronte di una alternativa bilaterale, con un un livello di significatività $\alpha = 0.05$. Infine, si valuti la bontà del modello calcolando il coefficiente di determinazione lineare.

A.40) Si vuole indagare se il consumo di alcool influisce sul consumo di tabacco. Si considera un campione di $n = 10$ soggetti e si ottengono misurazioni sulla variabile esplicativa x, che descrive la spesa settimanale media in sterline per gli alcolici, e sulla variabile risposta Y, che corrisponde la spesa settimanale media in sterline per tabacchi. Le osservazioni sono riportate nella seguente tabella

x	6.47	6.13	6.19	4.89	5.63	4.52	5.89	4.79	5.27	6.08
y	4.03	3.76	3.77	3.34	3.47	2.92	3.20	2.71	3.53	4.51

Nell'ipotesi che il modello di regressione lineare semplice sia appropriato, si determini la retta di regressione stimata. Con riferimento agli stimatori per l'intercetta e per il coefficiente angolare, si determinino gli standard error e si calcolino gli associati standard error stimati. Inoltre, si proponga una stima per la varianza di Y. Si verifichi l'ipotesi che il coefficiente angolare della retta di regressione sia nullo, a fronte di una alternativa bilaterale, con un un livello di significatività $\alpha = 0.01$. Infine, si valuti la bontà del modello calcolando il coefficiente di determinazione lineare.

A.41) Si sono osservati i seguenti $n = 15$ valori delle variabili esplicativa x e risposta Y

x	1.44	4.36	5.47	3.39	3.92	2.12	0.92	5.63
Y	10	798	816	42	155	51	3	832
x	5.76	2.72	2.71	3.62	4.05	1.63	0.24	
Y	1275	18	46	81	186	10	2	

Nell'ipotesi che il modello di regressione lineare semplice sia appropriato, si determini la retta di regressione stimata. Con riferimento agli stimatori per l'intercetta e per il coefficiente angolare, si determinino gli standard error e si calcolino gli associati standard error stimati. Inoltre, si proponga una stima per la varianza di Y. Si verifichi l'ipotesi che il coefficiente angolare della retta di regressione sia nullo, a fronte di una alternativa bilaterale, con un un livello di significatività $\alpha = 0.05$. Si valuti la bontà del modello calcolando il coefficiente di determinazione lineare. Infine, si stimi di nuovo il modello di regressione dopo aver trasformato i valori della variabile risposta con il logaritmo naturale. Si dica quale dei due modelli di regressione risulta migliore.

Alcuni degli esercizi proposti sono tratti dai seguenti testi.

Navidi, W. (2006). *Probabilità e statistica per l'ingegneria e le scienze.* McGraw-Hill.
Pauli, F., Torelli, N. e Trevisani, M. (2008). *Statistica: esercizi ed esempi.* Pearson.
Triola, M.M. e Triola, F.T. (2009). *Statistica per le discipline biosanitarie.* Pearson.

Soluzioni

Si forniscono le soluzioni di alcuni quesiti presenti negli esercizi proposti.

C.3) 0.4, 0.049.
C.4) 68.
C.6) 0.75, 0.25.
C.9) 0.03, 0.0054, $[0.019, 0.041]$, 247.
C.10) 1.8744, 0.18744, $[1.507, 2.242]$.
C.11) $[1.44, 4.44]$, $[1.15, 4.73]$, $[2.24, 3.64]$, $[2.10, 3.78]$.
C.12) $[1.43, 1.67]$, $[0.40, 0.64]$.
C.13) $[0.4981, 0.5014]$, $[0.4974, 0.5021]$, $[0.000047, 0.0002]$, $[0.000035, 0.0036]$.
C.14) $[1.25, 1.75]$.
C.15) $[0.359, 0.419]$.
C.16) $[3.65, 4.32]$.
C.17) $[1.25, 3.57]$.
C.18) 0.988, 0.12, 0.025.
C.19) Si rifiuta H_0, 0.0058.
C.20) Si rifiuta H_0, $2.34 \cdot 10^{-6}$.
C.21) 0.9508, 0.049, 0.098.
C.22) Si accetta H_0.
C.23) 0.06, 0.017, si rifiuta H_0.
C.24) Si accetta H_0.
C.25) Si accetta H_0.
C.26) Si rifiuta H_0.
C.27) Si rifiuta H_0.
C.28) Si accetta H_0.
C.29) Si accetta H_0.
C.30) Si accetta H_0, 0.226.
C.31) Si accetta H_0, 0.037.
C.32) Si accetta H_0, 0.063.
C.33) Si rifiuta H_0.
C.34) Si accetta H_0.
C.35) Si rifiuta H_0.
C.36) Si accetta H_0.
C.37) 0.816, si rifiuta H_0.
C.38) -0.512, si accetta H_0.
C.39) 4.028, -0.833, 2.954, 0.599, 1.183, si accetta H_0, 0.884.
C.40) 0.111, 0.611, 0.962, 0.171, 0.121, si rifiuta H_0, 0.615.
C.41) -346.39, 198.44, 93.40, 25.90, 62214.92, si rifiuta H_0, 0.669, 0.442, 1.169, 0.938.

Suggerimento: provare a risolvere gli esercizi utilizzando anche le funzioni di R.