STATISTICA APPLICATA

Prova scritta del 24 giugno 2019

1) La distribuzione di frequenza assoluta dell'età, in anni compiuti, 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	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. Inoltre, si confrontino le due distribuzioni di frequenza e si faccia un breve commento. Infine, nell'ipotesi di avere a disposizione i dati grezzi, si definiscano i comandi R necessari per calcolare la media aritmetica, la mediana, la varianza, gli indici di asimmetria e curtosi e per disegnare l'istogramma e il *boxplot*.

2) In una data popolazione il 10% degli individui è affetto da una determinata patologia. Per diagnosticare tale patologia si effettua un test ematico con le seguenti caratteristiche: il test risulta negativo per il 10% degli ammalati, mentre risulta positivo per il 20% dei sani. Se un individuo, scelto a caso dalla popolazione in esame, risulta positivo al test, quale è la probabilità che risulti effettivamente malato?

3) Si sono rilevati i livelli di conoscenza della Matematica e della Statistica con riferimento ad un gruppo di studenti. I risultati, con riferimento alla variabile "Livello di conoscenza della Matematica" e alla variabile "Livello di conoscenza della Statistica", sono riportati nella seguente tabella, con le rispettive frequenze assolute congiunte

	Statistica		
Matematica	sufficiente	buono	ottimo
sufficiente	15	14	9
buono	2	9	15
ottimo	3	6	18

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 utilizzando un opportuno test di ipotesi.

4) Si vuole stimare la frequenza relativa p di giovani che giudicano positivamente un nuovo programma televisivo. A tal fine si intervistano 700 giovani e si rileva che 306 di essi esprimono un giudizio positivo. Si definisca un modello statistico parametrico utile per descrivere l'esperimento in esame. Inoltre, 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 determini un intervallo di confidenza per p ad un livello di confidenza (approssimato) $1 - \alpha = 0.99$.

5) Si definisca il coefficiente di correlazione lineare, si evidenzi la sua utilità e si esprima una valutazione sui valori che esso può assumere. Infine, si ricordino le funzioni di R che permettono il calcolo del coefficiente di correlazione e dell'associato test di ipotesi.

6) Si elenchino gli assiomi del calcolo delle probabilità e si presenti almeno un risultato che si ottiene dalla loro applicazione.

STATISTICA APPLICATA

Prova scritta del 15 luglio 2019

1) La distribuzione di frequenza assoluta del numero di libri letti in un mese, con riferimento a 150 ragazzi di 10 anni, è riportata nella seguente tabella

No. di libri	0	1	2	3	4	5	Totale
No. di ragazzi	15	60	48	20	5	2	150

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. Infine, dopo aver spiegato come si possa ricostruire il vettore contenente i dati grezzi, si definiscano i comandi R necessari per calcolare la media aritmetica, la mediana, la varianza, gli indici di asimmetria e curtosi e per disegnare il diagramma a bastoncini.

2) 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, fornendo una motivazione basata sul calcolo delle associate probabilità finali.

3) Con riferimento ad un gruppo di 14 studenti universitari sono riportati, nella seguente tabella, i voti ottenuti nella prova scritta X e nella prova orale Y di un certo esame

Studente	X	Y	Studente	X	Y
1	29	22	8	23	25
2	29	28	9	21	24
3	27	30	10	21	24
4	24	25	11	19	25
5	24	27	12	19	24
6	23	23	13	18	25
7	23	22	14	18	23

Si calcolino le medie e le varianze marginali ed inoltre la covarianza e il coefficiente di correlazione lineare.

4) 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 μ e varianza σ^2, entrambe ignote. Si introduca un opportuno stimatore per μ e si determini l'associato *standard error*. Nell'ipotesi che, con riferimento ai risultati dell'indagine campionaria, si abbia $n = 50$, $\sum_{i=1}^{50} x_i = 94.15$ e $\sum_{i=1}^{50} x_i^2 = 249.61$, si fornisca una stima per μ e si calcoli l'associato *standard error* stimato.

Inoltre, sulla base del campione osservato, si verifichi l'ipotesi nulla $H_0 : \mu = 1.5$, al livello $\alpha = 0.05$, contro un'ipotesi alternativa unilaterale destra. Infine, si ricordi con quale funzione di R è possibile effettuare tale verifica di ipotesi e si indichino quali specifici argomenti devono essere utilizzati.

5) Si presenti il modello Poisson, fornendo la definizione e ricordando le principali proprietà. Inoltre, si evidenzino alcuni contesti sperimentali di applicazione. Infine, si ricordino le funzioni di R che permettono il calcolo della funzione di probabilità, della funzione di ripartizione, dei quantili e la simulazione di osservazioni.

6) Si definisca lo *standard error* e si discuta la sua utilità nell'ambito delle procedure di inferenza statistica.

STATISTICA APPLICATA

Prova scritta del 16 settembre 2019

1) La distribuzione di frequenza assoluta dei canoni di affitto annuo, in migliaia di euro, di 131 famiglie di una determinata area geografica è riportata nella seguente tabella

Canone	$8 \dashv 8.4$	$8.4 \dashv 8.8$	$8.8 \dashv 9.4$	$9.4 \dashv 9.8$	$9.8 \dashv 10$	$10 \dashv 10.2$	Totale
No. famiglie	25	27	19	12	15	33	131

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 le corrispondenti frequenze cumulate, assolute e relative. Si costruisca una rappresentazione grafica mediante l'istogramma. Inoltre, si individui la classe mediana e si calcoli la media aritmetica e la varianza di Y. Infine, nell'ipotesi di avere a disposizione i dati grezzi, si definiscano i comandi R necessari per calcolare la media aritmetica, la mediana, la varianza, gli indici di asimmetria e curtosi e per disegnare l'istogramma e il *boxplot*.

2) Una azienda utilizza tre impianti di produzione, indicati con A, B e C. Il 40% della produzione proviene da A, il 50% da B e il restante 10% da C. È noto che il 10% degli elementi prodotti da A è difettoso, mentre la percentuale degli elementi difettosi prodotti da B e C corrisponde rispettivamente al 5% e al 2%. Si sceglie a caso un elemento e si verifica che è difettoso. Si determini la probabilità che esso provenga dall'impianto A.

3) Con riferimento ad un gruppo di 14 studenti universitari sono riportati, nella seguente tabella, i voti ottenuti nella prova scritta X e nella prova orale Y di un certo esame

Studente	X	Y	Studente	X	Y
1	29	22	8	23	25
2	29	28	9	21	24
3	27	30	10	21	24
4	24	25	11	19	25
5	24	27	12	19	24
6	23	23	13	18	25
7	23	22	14	18	23

Si calcolino le medie e le varianze marginali ed inoltre la covarianza e il coefficiente di correlazione lineare.

4) Da indagini precedentemente svolte è noto che l'età alla quale si manifestano disturbi cardiovascolari per le donne può essere ragionevolmente descritta da una variabile casuale normale con media 48 anni. Si considera l'età di insorgenza di disturbi cardiovascolari per un campione di 17 donne, sia essa $x_1, \ldots, x_{17}$, e si ricava che $\sum_{i=1}^{17} x_i = 826.03$ e $\sum_{i=1}^{17} x_i^2 = 44065.6$. Utilizzando tali dati si vuole verificare l'ipotesi $H_0 : \mu = 48$, a fronte di un'alternativa bilaterale, considerando un livello di significatività $\alpha = 0.01$. Infine, si ricordi con quale funzione di R è possibile effettuare tale verifica di ipotesi e si indichino quali specifici argomenti devono essere utilizzati.

5) La nozione di quantile di livello α per una variabile casuale: definizione e procedure per il calcolo. Inoltre, si ricordino le funzioni di R che permettono il calcolo dei quantili di una variabile casuale con una determinata distribuzione di probabilità e si indichi come devono essere specificati i corrispondenti argomenti.

6) La variabile casuale varianza campionaria: definizione, principali proprietà e ambiti di applicazione nel contesto della statistica inferenziale.

STATISTICA APPLICATA

Prova scritta del 27 gennaio 2020

1) La distribuzione di frequenza assoluta del reddito familiare lordo dei residenti di una determinata città italiana è riportata nella seguente tabella, con i valori espressi in migliaia di euro,

Reddito	No. nuclei familiari
$0 \dashv 10$	100
$10 \dashv 20$	150
$20 \dashv 40$	350
$40 \dashv 50$	300
$50 \dashv 100$	100
Totale	1000

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 si calcoli la media aritmetica e la varianza di Y. Infine, nell'ipotesi di avere a disposizione i dati grezzi, si definiscano i comandi R necessari per calcolare la media aritmetica, la mediana, la varianza, gli indici di asimmetria e curtosi e per disegnare l'istogramma, il *boxplot* e la stima della densità.

2) Si considerino 8 urne, all'apparenza indistinguibili, che contengono 10 palline ciascuna. È noto che 4 urne contengono 7 palline bianche e 3 nere e le restanti 4 contengono 2 palline bianche e 8 nere. Viene scelta a caso un'urna e da essa si estraggono due palline senza reinserimento. Si determini la probabilità che le due palline siano bianche. Nell'ipotesi che le due palline estratte siano bianche, si determini la probabilità che si sia selezionata un'urna con 7 palline bianche.

3) Si sono rilevati i livelli di conoscenza della lingua inglese e della lingua francese con riferimento ad un gruppo di studenti. I risultati, con riferimento alla variabile "Livello di conoscenza dell'inglese" e alla variabile "Livello di conoscenza del francese", sono riportati nella seguente tabella, con le rispettive frequenze assolute congiunte

	Inglese		
Francese	sufficiente	buono	ottimo
sufficiente	15	14	9
buono	2	9	15
ottimo	3	6	18

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 utilizzando un opportuno test di ipotesi. Quale è la funzione di R con cui si può effettuare questa analisi?

4) Si vuole stimare la percentuale p di utenti che giudica positivamente il servizio fornito da un determinato call center. A tal fine si intervistano 1056 persone, scelte casualmente tra quelle che hanno utilizzato il servizio, delle quali 642 esprimono un giudizio positivo. Si definisca un modello statistico parametrico utile per descrivere l'esperimento in esame. Inoltre, si introduca un opportuno stimatore per p e si determini l'associato *standard error*. Inoltre, con riferimento ai risultati dell'indagine campionaria, si fornisca una stima per p, si calcoli l'associato *standard error* stimato e si ottenga un intervallo di confidenza per p di livello (approssimato) $1-\alpha = 0.95$. Infine, si indichino i comandi R necessari per calcolare stima, *standard error* stimato e intervallo di confidenza di livello 0.95.

5) Si definiscano l'istogramma e il boxplot e si evidenzi la loro utilità nel contesto delle analisi esplorative. Infine, si ricordino le funzioni di R che permettono la costruzione di tali rappresentazioni grafiche.

6) La variabile casuale media campionaria: definizione, principali proprietà e applicazioni nell'ambito della statistica inferenziale.

STATISTICA APPLICATA

Prova scritta del 17 febbraio 2020

1) La distribuzione di frequenza assoluta del numero di figli rilevato con riferimento a 150 famiglie è riportata nella seguente tabella

No. di figli	0	1	2	3	4	5	Totale
No. nuclei famigliari	25	45	50	15	10	5	150

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. Infine, nell'ipotesi di avere a disposizione i dati grezzi, si definiscano i comandi R necessari per calcolare la media aritmetica, la mediana, la varianza, gli indici di asimmetria e curtosi e per produrre una opportuna rappresentazione grafica per la distribuzione di frequenza.

2) Si consideri un'urna contenente 10 palline, delle quali 3 sono nere. Si lanci una moneta regolare e quindi si estragga dall'urna un campione, senza reinserimento, di ampiezza pari a 1, se la moneta ha dato croce, o pari a 2, se la moneta ha dato testa. Si determini la probabilità che tutte le palline del campione siano nere. Se tutte le palline del campione sono nere (e non si conosce l'esito del lancio della moneta), quale è la probabilità che la moneta abbia dato croce?

3) Si è misurata la temperatura al suolo in gradi centigradi Y e la velocità del vento in m al secondo X per otto volte a distanza di un'ora. I dati sono riportati nella seguente tabella

Misurazione	X	Y
1	2.2	38.6
2	2.4	39.3
3	2.5	39.7
4	2.3	38.9
5	2.1	38.3
6	1.8	37.3
7	0.9	34.3
8	1.2	35.3

Si calcolino le medie e le varianze marginali ed inoltre la covarianza e il coefficiente di correlazione lineare. Quali sono i comandi di R con cui si possono ottenere questi risultati? Infine, con quale funzione di R è possibile svolgere l'associato test di ipotesi sul coefficiente di correlazione lineare?

4) 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 il valore atteso $\mu = 1/\lambda$ 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) $1 - \alpha = 0.95$. Infine, si indichino i comandi R necessari per calcolare stima, *standard error* stimato e intervallo di confidenza di livello 0.95.

5) La nozione di mediana per una variabile statistica: definizione e procedure per il calcolo. Si ricordi anche la funzione di R che fornisce il valore della mediana.

6) Dato un campione casuale costituito da variabili casuali $X_1, \ldots, X_n$, $n \geq 1$, indipendenti con distribuzione di probabilità normale con media e varianza ignote, si definisca la regione di rifiuto di un test bilaterale per $H_0 : \mu = \mu_0$ di livello α. Infine, si ricordi la funzione di R che esegue tale test.