

Esercitazioni di Statistica

1 - Distribuzioni di frequenza

Esercizio 1 Scaricare il file [Customers.xls](#). Stabilire la tipologia dei caratteri presenti nel dataset.

Per i prossimi esercizi si considera la popolazione costituita dai primi 10 clienti presenti nel file Customers. I risultati saranno ottenuti senza l'ausilio del foglio di calcolo.

Esercizio 2 Con riferimento al carattere **Figli**:

1. indicare la natura del carattere;
2. indicare la tipologia di distribuzione dei dati;
3. indicare le unità statistiche;
4. determinare il numero di unità statistiche e di modalità rilevate;
5. ottenere le distribuzioni di frequenze assolute, relative e percentuali;
6. ottenere la distribuzione di frequenza assoluta, suddividendo le osservazioni nelle classi $[0, 1]$, $[2]$, $[3, 4]$.

Esercizio 3 Con riferimento ai caratteri **Figli** e **Genere**:

1. indicare la natura dei caratteri;
2. indicare la tipologia di distribuzione dei dati;
3. indicare le unità statistiche;

4. determinare il numero di unità statistiche, il numero di modalità per ciascun carattere ed il numero di coppie di modalità;
5. costruire le distribuzioni di frequenze assolute, congiunte e marginali.

Per i prossimi esercizi si considera la popolazione costituita da tutti i clienti presenti nel file Customers. Data l'ampiezza della popolazione, i risultati saranno ottenuti utilizzando le funzioni e gli strumenti di un foglio di calcolo.

Esercizio 4 Utilizzando il carattere **Num.acquisti**,

1. indicare la tipologia di distribuzione dei dati;
2. indicare le unità statistiche e il loro numero;
3. costruire la distribuzione di frequenza determinando le frequenze assolute, relative e percentuali;
4. determinare il numero di modalità del carattere.

Suggerimento: sul file [istruzioni-excel.pdf](#) sono elencate diverse funzioni e strumenti utili per lavorare con i fogli di calcolo.

Esercizio 5 Utilizzando il carattere **Distanza**

1. indicare la natura del carattere;
2. costruire la distribuzione di frequenze assolute e commentare il risultato ottenuto.

Esercizio 6 Utilizzando i caratteri **Num.acquisti** e **Figli** del file Customers

1. indicare il numero di modalità delle distribuzioni marginali dei due caratteri;
2. determinare il numero di coppie di modalità;
3. costruire la distribuzione congiunta di frequenza determinando le frequenze congiunte assolute e le frequenze assolute marginali.

Esercizio 7 Utilizzando il caratteri **Val.acquisti** del file Customers,

1. ottenere la distribuzione di frequenza assoluta, suddividendo le osservazioni nelle classi $[0, 100)$, $[100, 150)$, $[150, 200)$, $[200, 250)$, $[250, 400)$;
2. costruire l'istogramma.

Esercizio 8* Aggiungere al file Customers l'indicazione della provincia di residenza di ogni cliente.

Suggerimento: trovare una base dati contenente l'informazione relativa alla provincia di appartenenza di ogni comune. Se non si riesce a trovare autonomamente una base dati, se ne può trovare una al link

<http://lab.comuni-italiani.it/download/comuni.html>.

Usare poi la funzione CERCA.VERT di Excel per unire le informazioni dalle due basi di dati.

I prossimi esercizi non sono basati su file di dati.

Esercizio 9 Stabilire la tipologia dei seguenti caratteri:

Numero di componenti della famiglia	Prezzo di un chilo di pane in Euro
Reddito familiare	Voto all'esame di statistica
Salario in Euro	Marca di caffè preferita
Classe di reddito (I, II, ecc.)	Consumo giornaliero di calorie
Ore di studio al giorno	Regione di provenienza
Tempo dedicato allo studio al giorno	Colore degli occhi
Livello di inquadramento lavorativo	Sistema operativo (di un computer)
Attività lavorativa	Tasso di interesse annuale
Grado di soddisfazione di un cliente	Numero di scarpa

Esercizio 10 Sia X il carattere che indica il numero di guasti mensili di un macchinario. Avendo rilevato il numero di guasti nell'anno 2015, si ottiene la seguente distribuzione:

2 1 2 1 3 2 0 2 0 1 2 2

1. Indicare la tipologia di distribuzione dei dati;
2. indicare le unità statistiche;
3. indicare la natura del carattere;
4. determinare il numero di unità statistiche e di modalità;
5. costruire la distribuzione di frequenza determinando le frequenze assolute, relative e percentuali.

Esercizio 11 Sui prodotti di un lotto di produzione, sono stati rilevati i caratteri $X_1 = \{\text{Funzionalità prodotto}\}$ e $X_2 = \{\text{Processo produttivo}\}$. Il carattere X_1 assume le modalità difettoso e non difettoso, mentre X_2 assume le modalità 1, 2 e 3, a seconda del macchinario utilizzato. Si ottiene la seguente distribuzione

X_1	X_2
difettoso	1
difettoso	2
non difettoso	1
difettoso	3
non difettoso	3
non difettoso	2
difettoso	2
non difettoso	3
difettoso	1
non difettoso	1
non difettoso	2
non difettoso	1
difettoso	3
difettoso	3

1. Indicare la tipologia di distribuzione dei dati;
2. indicare le unità statistiche;
3. indicare la natura dei caratteri;
4. determinare il numero di unità statistiche e di coppie di modalità;
5. indicare il numero di modalità delle distribuzioni marginali dei due caratteri;
6. costruire le distribuzioni di frequenze assolute congiunte e marginali dei due caratteri.

Esercizio 12 La seguente tabella riporta il numero di impiegati, per settore (X) e genere (Y), in un certo comune.

X	Y	
	M	F
Agricoltura	140	76
Industria	730	160
Servizi	510	470

1. indicare la tipologia di distribuzione dei dati;
2. indicare la popolazione e le unità statistiche;
3. ottenere le distribuzioni di frequenze assolute marginali;
4. ottenere la distribuzione di frequenze relative congiunte;
5. ottenere le distribuzioni di frequenze relative marginali.

Esercizio 13 In un'indagine sui consumi delle auto a benzina nei percorsi urbani è stata osservata la distribuzione del numero di litri consumati per 100 Km riportata nella seguente tabella:

Consumi	Frequenza
5 † 10	15
10 † 15	45
15 † 25	35
25 † 35	4
35 † 38	1

1. indicare la tipologia di distribuzione dei dati;
2. costruire l'istogramma, individuare la classe modale e commentare i risultati.

Esercizio 14 Su un collettivo formato da 120 maschi e 80 femmine è stata rilevata l'età in anni compiuti, ottenendo la seguente distribuzione percentuale per genere:

Età	% Maschi	% Femmine
0 † 20	10	20
20 † 30	10	20
30 † 50	30	30
50 † 90	50	30
	100	100

1. indicare la tipologia di distribuzione dei dati;
2. calcolare il numero di unità statistiche nel collettivo di età inferiore ai 20 anni;
3. calcolare la percentuale di unità statistiche nel collettivo di età maggiore o uguale ai 50 anni;
4. calcolare il numero di maschi di età maggiore o uguale ai 30 anni.

Esercizio 15 In un collettivo di pazienti sono stati rilevati la quantità di colesterolo in mg per 100 ml di sangue e il genere

Colesterolo	Genere	
	M	F
120 † 160	40	20
160 † 180	10	12
180 † 200	20	10
200 † 240	10	20
240 † 300	45	10

1. indicare la tipologia di distribuzione dei dati;
2. rappresentare graficamente la distribuzione del colesterolo nell'intera popolazione (marginalmente rispetto al genere) e commentare i risultati.

Esercizio 16 Con riferimento all'esercizio 5, indicare una possibile rappresentazione sintetica dei dati.

I prossimi esercizi si basano su dataset disponibili su internet.

Esercizio 17 Sul rapporto [Italia in cifre 2016](#) sono disponibili dati su diversi aspetti economici, demografici, sociali e ambientali dell'Italia.

1. Si consideri la prima tabella della sezione "Condizioni di vita". Indicare
 - la popolazione e le unità statistiche;
 - i caratteri considerati e la loro natura;
 - la tipologia di distribuzione dei dati.
2. ottenere il numero di donne occupate al Nord nell'anno 2015;
3. ottenere un istogramma della percentuale della superficie comunale dei capoluoghi di regione che nel 2014 era riservata al verde urbano. Suddividere i dati in 6 classi.

Esercizio 18 Un'azienda vuole valutare la possibilità di lanciare sul mercato una nuova bibita in lattina. A tale scopo incarica il settore Marketing di effettuare un'indagine sulla propensione all'acquisto di una nuova bibita e sulle relative caratteristiche. Il settore Marketing decide di raccogliere le informazioni attraverso la predisposizione e la somministrazione a 36 soggetti di un questionario.

I dati ottenuti sono contenuti nel file [datibibita.xls](#).

Usando un foglio di calcolo, ottenere la distribuzione di frequenza della variabile **Ipocaloricità** con l'istruzione FREQUENZA e di **Sesso** utilizzando lo strumento tabella pivot. Ottenere inoltre la distribuzione congiunta di frequenze per i due caratteri. Il risultato finale dovrebbe essere simile, nella forma, a quanto contenuto nel file [vuoto.xls](#).

Esercizio 19 Scaricare dal sito <http://noi-italia.istat.it/> (edizione 2018) i dati relativi alle Persone di 3 anni e più che praticano sport, qualche attività fisica e persone sedentarie per regione, Anno 2017 (valori percentuali).

È possibile reperire il dataset scegliendo Popolazione e Società -> Cultura e tempo libero -> Dati. Si aprirà una cartella zip, contenente due file. Aprire quello denominato "Cultura e tempo libero_altri dati" e posizionarsi sul foglio di lavoro denominato "Pratica sport - Regione".

1. Verificare che la somma delle frequenze percentuali per l'Italia sia 100;
2. Rappresentare graficamente la distribuzione di frequenza della pratica di attività fisica per l'Italia;
3. Rappresentare la distribuzione di frequenza della pratica di attività fisica per la regione Lazio utilizzando la stessa scala, stesso minimo e massimo per le frequenze percentuali.