

RAPPRESENTAZIONI GRAFICHE DI FENOMENI STATISTICI

Laboratorio di Matematica Computazionale

Massimo Paschetto

I.T.S. “Cangrande della Scala” – Verona

25 settembre 2019

La *statistica* [1] è la scienza che studia i fenomeni collettivi. In uno studio statistico si parla di *popolazione* in riferimento all'insieme di elementi studiati ciascuno dei quali si chiama individuo o *unità statistica*. Di ogni unità statistica si studiano i *caratteri*, cioè le caratteristiche che si intendono indagare. I caratteri possono essere:

- *qualitativi* come sesso, colore dei capelli, l'essere o meno iscritti all'Università, titolo di studio o
- *quantitativi*, cioè esprimono delle quantità misurabili, come peso, altezza, reddito.

I caratteri qualitativi si possono dividere in:

- **qualitativi sconnessi**, non ordinabili come per esempio il sesso;
- **qualitativi connessi**, ordinabili come per esempio il titolo di studio.

I caratteri quantitativi al loro volta possono essere divisi in:

- **quantitativi discreti**, si rappresentano utilizzando insiemi numerabili, solitamente i numeri naturali;
- **quantitativi continui**, si esprimono utilizzando insiemi continui, solitamente i numeri reali.

Se i caratteri, qualitativi o quantitativi, vengono codificati con opportuni insiemi numerici si parla di *variabile statistica*.

La *statistica descrittiva* o *analisi dei dati* si occupa di evidenziare attraverso opportune sintesi numeriche e grafiche le caratteristiche di opportuni fenomeni collettivi.

Le principali fasi dell'indagine statistica sono:

1. Pianificazione e rilevazione
2. Elaborazione
3. Presentazione
4. Interpretazione

La fase di presentazione consiste nell'esposizione dei dati opportunamente sintetizzati in tabelle, grafici, indici, funzioni. In questo documento vedremo come possiamo utilizzare Python per creare eleganti grafici di sintesi a partire da dati grezzi o tabelle.

Definire i concetti di distribuzione statistica, valori rilevati, frequenze assolute, frequenze relative, frequenze cumulate, funzione di ripartizione.

1 Rappresentazioni grafiche

Le rappresentazioni grafiche servono per facilitare la visualizzazione della composizione di un insieme di dati, cioè della variabile statistica.

I vantaggi delle rappresentazioni grafiche sono fornire una visione sintetica e facilitare l'interpretazione, l'inconveniente è di mancare di precisione e soprattutto di essere soggettive, cioè di permettere letture diverse degli stessi dati.

Il giudizio su una rappresentazione grafica si può basare su 5 aspetti:

1. Accuratezza: precisione nei dettagli
2. Semplicità: uso di soli elementi grafici.

3. Chiarezza: capacità di comunicare senza ambiguità
4. Aspetto: è necessario che sia il più possibile armonioso.
5. Struttura: ben definita.

I principali tipi di rappresentazioni grafiche sono:

- **Diagrammi a barre**, per caratteri qualitativi ordinati o caratteri quantitativi discreti;
- **Diagrammi a nastri**, per caratteri qualitativi non ordinati;
- **Diagrammi a aree**, per caratteri quantitativi continui nel tempo;
- **Istogrammi**, per caratteri quantitativi continui suddivisi in classi;
- **Diagrammi a settori circolari**, per caratteri qualitativi non ordinati o ordinati ciclici;
- **Diagrammi polari**, per caratteri ciclici;
- **Cartogrammi**, per serie territoriali;
- **Ideogrammi**, per caratteri qualitativi e quantitativi;
- **Diagrammi cartesiani**, per serie temporali.

2 Diagrammi a barre

I grafici a barre sono uno dei metodi più comuni per visualizzare i dati. Il confronto delle informazioni è immediato, i valori massimi e minimi dei dati si notano a colpo d'occhio. Si prestano bene per la comparazione di dati tra più categorie.

Selezionare il modello a barre affiancate o sovrapposte e assegnare un colore diverso a ogni categoria, crea un maggiore impatto visivo.

I diagrammi a barre o a colonne (*bar chart*) sono tra i più utilizzati per rappresentare dati e variabili discrete. Per rappresentare una tabella si disegna, per ciascuna modalità del fenomeno, un rettangolo di altezza proporzionale alla frequenza o alla intensità della modalità e di larghezza arbitraria ma costante

I rettangoli possono essere sia verticali (diagramma a colonne fig. 2) sia orizzontali (diagramma a nastri).

Grafici a barre Nei diagrammi a barre e a nastri (ortogrammi) ogni frequenza o intensità della distribuzione viene rappresentata da una barra o da un nastro in modo da ottenere una successione di rettangoli con la stessa base (o altezza) e le altezza (o le basi) proporzionali alle frequenze o quantità. Quando il carattere è qualitativo ordinato o quantitativo, è preferibile utilizzare il grafico a barre poiché le barre poste sull'asse orizzontale permettono di cogliere meglio l'ordinamento delle modalità.

INVALSI 2012 Un gruppo di boy scout è formato da ragazze e ragazzi di età compresa tra i 10 e i 14 anni. La distribuzione delle frequenze percentuali delle età è riportata nel diagramma seguente.

Sulla base dei dati riportati nel diagramma, indicare se ciascuna delle seguenti affermazioni è vera (V) o falsa (F).

- (a) Più dell'80% dei ragazzi ha meno di 13 anni.
- (b) Meno del 70% dei ragazzi ha più di 11 anni.
- (c) La percentuale dei ragazzi che hanno 12 o 14 anni è uguale alla percentuale di ragazzi che hanno 10 o 11 o 13 anni.

V	F
V	F
V	F

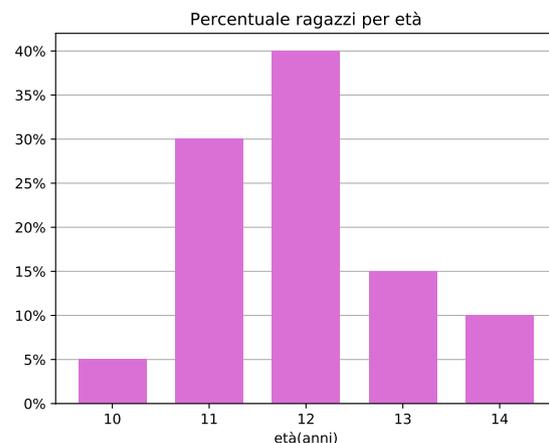


Figura 1 – Diagramma a barre (ortogramma)

```

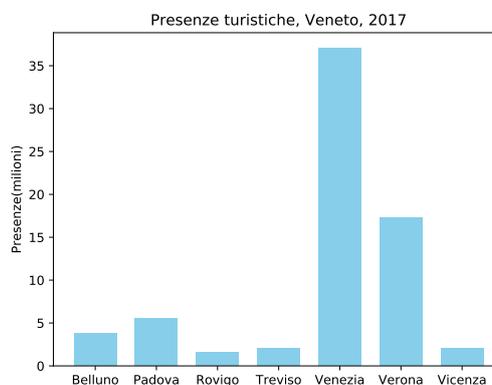
1 import numpy as np
2 import matplotlib.pyplot as plt
3
4 x = [10, 11, 12, 13, 14]
5 y = [5, 30, 40, 15, 10]
6 widthBar = 0.70
7
8 fig, ax = plt.subplots(1, 1)
9
10 rect = ax.bar(x, y, width=widthBar, color='Orchid')
11 ax.set_axisbelow(True)
12 griglia = ax.grid(True, axis='y', linestyle='-')
13
14 ax.set_xticks(x)
15 ax.set_xlabel('età(anni)')
16 yTicks = [0, 5, 10, 15, 20, 25, 30, 35, 40]
17 yLabels = ['0%', '5%', '10%', '15%', '20%', '25%', '30%', '35%', '40%']
18 ax.set_yticks(yTicks)
19 ax.set_yticklabels(yLabels)
20 ax.set_title('Percentuale ragazzi per età')
21
22 plt.show()

```

Programma 1 – Diagramma a barre figura 1

Provincia	Presenze (milioni)
Belluno	3,8
Padova	5,5
Rovigo	1,6
Treviso	2,0
Venezia	37,0
Verona	17,3
Vicenza	2,0
TOTALE	69,2

(a) Tabella dati



(b) Diagramma a barre (Bar Chart)

Figura 2 – Il turismo nelle province venete. Anno 2017. Fonte: Elaborazioni dell'Ufficio di Statistica della Regione del Veneto su dati provvisori Istat - Regione Veneto, <http://www.regione.veneto.it/web/statistica>

```

1 import numpy as np
2 import matplotlib.pyplot as plt
3
4 labels = ['Belluno', 'Padova', 'Rovigo', 'Treviso', 'Venezia', 'Verona', 'Vicenza']
5 x = [1, 2, 3, 4, 5, 6, 7]
6 y = [3.8, 5.5, 1.6, 2.0, 37.0, 17.3, 2.0]
7 widthBar = 0.70
8
9 fig, ax = plt.subplots(1, 1)
10
11 rect = ax.bar(x, y, width=widthBar, color='SkyBlue')
12 ax.set_xticks(x)
13 ax.set_xticklabels(labels)
14 ax.set_ylabel('Presenze(milioni)')
15 ax.set_title('Presenze turistiche, Veneto, 2017')
16
17 plt.show()

```

Programma 2 – Diagramma a barre figura 2

I grafici a nastri, così come quelli a barre, sono particolarmente adatti a rappresentare caratteri quantitativi discreti quali per esempio: n. di componenti del nucleo familiare o voto ad un certo esame.

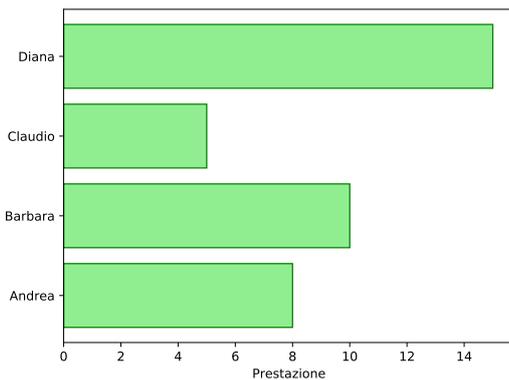


Figura 3 – Diagramma a nastri

```

1 import numpy as np
2 import matplotlib.pyplot as plt
3
4 y_pos = [0, 1, 2, 3]
5 nomi = ['Andrea', 'Barbara', 'Claudio', 'Diana']
6 x = [8, 10, 5, 15]
7
8 fig, ax = plt.subplots(1, 1)
9 rects = ax.barh(y_pos, x, color='lightgreen',
10 edgecolor='green')
11 ax.set_yticks(y_pos)
12 ax.set_yticklabels(nomi)
13 ax.set_xlabel('Prestazione')
14 plt.show()

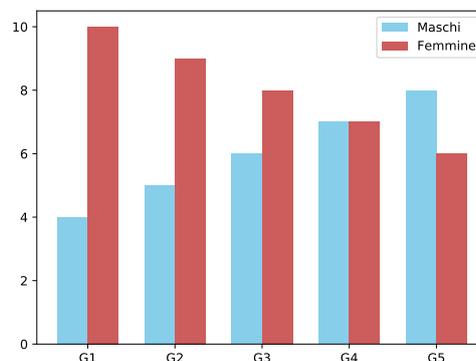
```

Programma 3 – Diagramma a nastri

Quando per uno stesso carattere si osservano due o più distribuzioni semplici, relativamente a diversi collettivi, possiamo metterle a confronto riportandole in un unico grafico con i cosiddetti grafici a barre o a nastri multipli.

Dato rilevato	Maschi	Femmine
G1	4	10
G2	5	9
G3	6	8
G4	7	7
G5	8	6
TOTALE	30	40

(a) *Tabella dati*



(b) *Diagramma a due colonne affiancate*

Figura 4 – Diagramma a due colonne(maschi/femmine)

```

1 import numpy as np
2 import matplotlib.pyplot as plt
3
4 y1 = [4,5,6,7,8]
5 y2 = [10,9,8,7,6]
6 x = np.array([0, 1, 2, 3, 4])
7 widthBar = 0.35
8
9 fig, ax = plt.subplots(1, 1)
10
11 rects1 = ax.bar(x - widthBar/2, y1, width=widthBar, color='SkyBlue', label='Maschi')
12 rects2 = ax.bar(x + widthBar/2, y2, width=widthBar, color='IndianRed', label='Femmine')
13 ax.set_xticks(x)
14 ax.set_xticklabels(('G1', 'G2', 'G3', 'G4', 'G5'))
15 ax.legend()
16
17 plt.show()

```

Programma 4 – Diagramma a due colonne affiancate figura 4

3 Areogrammi

L'areogramma è un tipo di rappresentazione grafica di dati statistici in cui le frequenze vengono rappresentate da aree di opportune figure geometriche piane, proporzionali alla distribuzione delle frequenze.

Fanno parte della categoria degli areogrammi i diagrammi a settori circolari, gli istogrammi e i

I grafici a colonne, invece, non sono areogrammi perchè la frequenza che rappresentano corrisponde all'altezza dei rettangoli e non all'area.

3.1 Diagrammi a settori circolari

I grafici a settori circolari (*pie chart*) sono areogrammi utili quando si vuole evidenziare come il fenomeno sia suddiviso fra le varie modalità che lo compongono. La frequenza o l'intensità delle modalità rilevate è rappresentata dall'area dei settori circolari. La totalità del fenomeno è rappresentata dall'intera circonferenza. L'angolo al centro di ciascun settore è proporzionale alla frequenza (o intensità) della modalità che rappresenta il settore circolare. Le ampiezze α_i degli angoli al centro dei settori circolari si calcolano utilizzando la seguente proporzione

Ampiezza settori circolari

$$\alpha_i : 360 = f_i : f_T$$

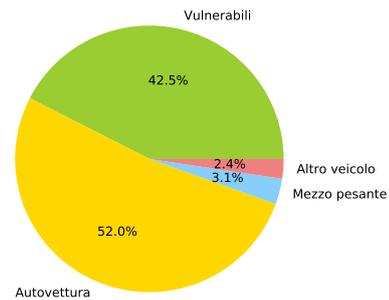
dove f_i è la frequenza della i -esima modalità e f_T è il totale delle frequenze.

Categoria	Freq. assoluta	Freq. relativa
Vulnerabili*	1729	0,425
Autovettura	2117	0,520
Mezzo pesante	127	0,031
Altro veicolo	97	0,024
TOTALE	4070	1,000

*: pedoni, ciclisti, motociclisti

(a) Tabella dati

Utenti della strada morti o feriti in Veneto, 2017



(b) Diagramma a settori circolari (Pie Chart)

Figura 5 – Utenti della strada morti o feriti per categoria di utenza, provincia di Verona, 2017. Fonte: Ufficio di Statistica della Regione del Veneto - Regione Veneto, <http://www.regione.veneto.it/web/statistica>

Nel codice Python che segue vengono definite tre liste contenenti rispettivamente la frequenza assoluta delle unità statistiche rilevate, le etichette che le definiscono e il colore da assegnare ad ogni settore circolare. Si osservi che nella riga 7, in cui viene definita la figura e gli assi che conterranno il diagramma, è consigliabile inserire l'impostazione `subplot_kw=dict(aspect='equal')` che impone la stessa unità di misura per essere sicuri che il diagramma a settori circolari non venga visualizzato come una ellisse.

Il comando della riga 9 consente di realizzare il diagramma a settori circolari inserendo automaticamente le percentuali di ciascun settore grazie all'opzione `autopct='%1.1f%%'` che approssima le percentuali ai decimi.

Le rimanenti istruzioni sono intuitive .

```

1 import numpy as np
2 import matplotlib.pyplot as plt
3
4 freq = [1729, 2117, 127, 97]
5 labels = ['Vulnerabili', 'Autovettura', 'Mezzo pesante', 'Altro veicolo']
6 colors = ['yellowgreen', 'gold', 'lightskyblue', 'lightcoral']
7 fig, ax = plt.subplots(1, 1, subplot_kw=dict(aspect='equal'))
8
9 torta = ax.pie(freq, colors=colors, labels=labels, explode=[0,0,0,0],
10               autopct='%1.1f%%', shadow=False, startangle=0)
11
12 ax.set_title('Utenti della strada morti o feriti in Veneto, 2017')
13
14 plt.show()

```

Programma 5 – Diagramma a settori circolari figura 5

3.2 Istogrammi

I grafici a colonne vengono utilizzati prevalentemente per rappresentare distribuzioni con mutabili statistiche discrete; per rappresentare le distribuzioni delle variabili continue, raggruppate per classi di valori, si utilizzano più correttamente gli istogrammi.

Asse y: densità di frequenza

Gli istogrammi rientrano tra i metodi che utilizzano le aree per rappresentare le distribuzioni.

Gli istogrammi, come i grafici a colonne, sono costituiti da tanti rettangoli quante sono le classi da rappresentare, ma a differenza dei grafici a colonne in cui i rettangoli sono per lo più distanziati, negli istogrammi i rettangoli sono contigui per indicare la continuità del fenomeno rilevato.

Grafici per variabili continue: l'istogramma • È utilizzato quanto la distribuzione si riferisce ad un carattere quantitativo continuo. • In un sistema di assi cartesiani è composto da una serie di rettangoli che hanno come base l'ampiezza delle varie classi in cui è stata ripartita la variabile d'interesse, e come altezza le frequenze

(assolute, relative semplici o percentuali). • Le basi dei rettangoli sono uguali; di conseguenza, le altezze sono proporzionali alle frequenze. È indifferente ragionare in termini di altezze o di aree di ogni rettangolo.

Istogrammi (... se le classi hanno la stessa ampiezza) Si rappresentino i dati della tabella successiva riferiti alla classificazione di 110 oggetti secondo il peso.

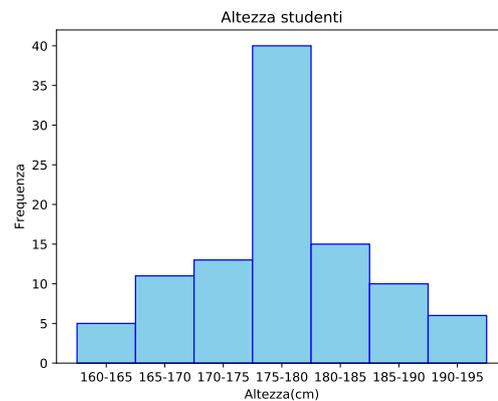
Istogrammi (... se le classi hanno ampiezza differente) Se le classi sono di ampiezza diversa, le frequenze non sono direttamente confrontabili ed è necessario rendere l'altezza proporzionale. Tale proporzione, è un rapporto tra la frequenza e l'ampiezza (indicata con D_i) di una classe, è definita densità di frequenza. Le densità di frequenza sono fra loro confrontabili. La densità di frequenza è assoluta o relativa a seconda del tipo di frequenza utilizzato nel calcolo. In un istogramma di frequenza ad ogni classe è associato un rettangolo con:

- base pari all'ampiezza di classe;
- l'altezza uguale alla densità di frequenza;
- l'area, per costruzione, pari alla frequenza associata alla classe.

Es. Istogrammi (... se le classi hanno ampiezza diversa) La seguente tabella riporta la distribuzione della popolazione residente nella regione Umbria per fasce d'età al 1 gennaio 2001 (fonte ISTAT)

Altezza(cm)	Frequenza
160 – 165	5
165 – 170	11
170 – 175	13
175 – 180	40
180 – 185	15
185 – 190	10
190 – 195	6
TOTALE	100

(a) Tabella dati



(b) Istogramma con classi di uguale ampiezza

Figura 6 – Altezza studenti

Le basi del rettangolo rappresentano le ampiezze delle classi, mentre le altezze rappresentano le frequenze delle classi. Se le classi sono di uguale ampiezza, i rettangoli hanno le basi uguali tra loro e le altezze sono proporzionali alle frequenze, come in figura 6.

```

1 import numpy as np
2 import matplotlib.pyplot as plt
3
4 freq = (5, 11, 13, 40, 15, 10, 6)
5 x = (162.5, 167.5, 172.5, 177.5, 182.5, 187.5, 192.5)
6 labels = ('160-165', '165-170', '170-175', '175-180', '180-185', '185-190', '190-195')
7 widthBar = 5
8
9 fig, ax = plt.subplots(1, 1)
10
11 rect = ax.bar(x, freq, widthBar, color='SkyBlue', edgecolor='Blue', label='Altezze')
12 # Aggiungo titolo, etichette per l'asse orizzontale e verticale, etc.
13 ax.set_title('Altezza studenti')
14 ax.set_xticks(x)
15 ax.set_xticklabels(labels)
16 ax.set_ylabel('Frequenza')
17 ax.set_xlabel('Altezza (cm)')
18
19 plt.show()

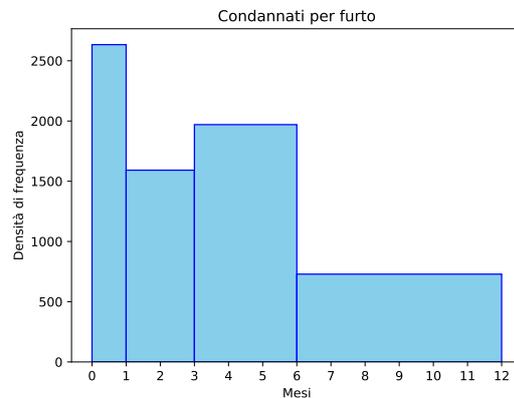
```

Programma 6 – Istogramma con classi di uguale ampiezza figura 6

Se le classi sono di ampiezza diverse, bisogna tenere conto che le altezze dei rettangoli devono essere proporzionali alla densità di frequenza, ovvero alla frequenza divisa per l'ampiezza della classe, per poter mantenere la proporzionalità tra le aree, si veda figura 7.

Mesi di reclusione	Numero di condannati	Ampiezza della classe	Densità di frequenza
fino a 1	2634	1	2634
1 – 3	3184	2	1592
3 – 6	5910	3	1970
6 – 12	4374	6	729
TOTALE	16102		

(a) Tabella dati



(b) Istogramma con classi di diversa ampiezza

Figura 7 – Condannati per furto nel 1984 (condanne non superiori a 12 mesi)

```

1 import numpy as np
2 import matplotlib.pyplot as plt
3
4 densFreq = (2634, 1592, 1970, 729)
5 x = (0.5, 2, 4.5, 9)
6 xTicks = (0,1,2,3,4,5,6,7,8,9,10,11,12)
7 widthBar = (1, 2, 3, 6)
8
9 fig, ax = plt.subplots(1, 1)
10
11 rect = ax.bar(x, densFreq, widthBar, color='SkyBlue', edgecolor='Blue')
12 ax.set_title('Condannati per furto')
13 ax.set_xticks(xTicks)
14 ax.set_ylabel('Densità di frequenza')
15 ax.set_xlabel('Mesi')
16
17 plt.show()

```

Programma 7 – Istogramma classi di diversa ampiezza

4 Diagrammi polari

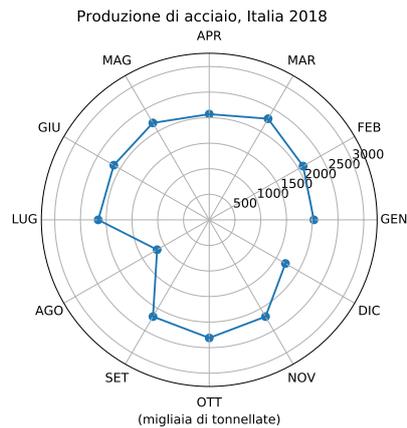
I diagrammi polari, costruiti in coordinate polari, sono utilizzati con le serie storiche e in particolare con le mutabili cicliche come ad esempio i giorni della settimana o i mesi dell'anno.

Per costruire un diagramma in coordinate polari si deve:

- dividere, mediante raggi vettore, l'angolo giro in tanti settori circolari di ampiezza costante quante sono le modalità della serie;
- assumere, a partire dal centro (polo), su ciascun raggio vettore un segmento di lunghezza proporzionale alla frequenza della modalità rappresentata da quel raggio.

Infine si congiungono generalmente i punti con una linea spezzata. Bisogna però osservare che non è del tutto corretto utilizzare delle linee rette ma si dovrebbero utilizzare degli archi di circonferenza.

mesi	tonn.(migliaia)
Gennaio	2031
Febbraio	2104
Marzo	2285
Aprile	2069
Maggio	2190
Giugno	2140
Luglio	2156
Agosto	1168
Settembre	2185
Ottobre	2310
Novembre	2186
Dicembre	1708
TOTALE	24532

(a) *Tabella dati*(b) *Diagramma polare***Figura 8** – Produzione di acciaio, Italia 2018, <http://www.federacciai.it>

Codice Python per realizzare il diagramma polare.

```

1 import numpy as np
2 import matplotlib.pyplot as plt
3
4 N = 12
5 etichette = ['GEN', 'FEB', 'MAR', 'APR', 'MAG', 'GIU',
6             'LUG', 'AGO', 'SET', 'OTT', 'NOV', 'DIC']
7 x = [2031, 2104, 2285, 2069, 2190, 2140, 2156, 1168, 2185, 2310, 2186, 1708]
8 theta = np.linspace(0.0, 2 * np.pi, N+1)
9
10 fig, ax = plt.subplots(1, 1, subplot_kw=dict(polar=True))
11
12 punti = ax.scatter(theta[:-1], x)
13 linea = ax.plot(theta[:-1], x)
14 fig.suptitle('Produzione di acciaio, Italia 2018', fontsize=12)
15 ax.set_xticks(theta[:-1])
16 ax.set_xticklabels(etichette)
17 ax.set_xlabel('(migliaia di tonnellate)', position=(0.5,0.5))
18
19 plt.show()

```

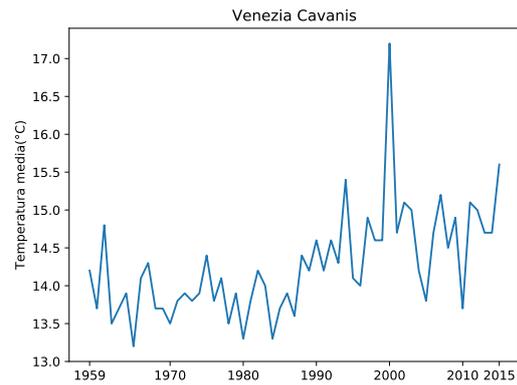
Programma 8 – Esempio di codice Python

5 Diagrammi cartesiani

Si utilizzano per rappresentare le serie storiche, soprattutto più serie congiuntamente. È un grafico costituito da una serie di punti individuati su un piano cartesiano, in cui sull'asse delle ascisse è posto il tempo e su quello delle ordinate il carattere osservato. I punti tracciati sul piano vengono uniti da segmenti che nel loro insieme costituiscono una spezzata che rappresenta l'andamento del fenomeno.

Anno	Temp	Anno	Temp	Anno	Temp	Anno	Temp
1959	14,2	1974	13,9	1989	14,2	2004	14,2
1960	13,7	1975	14,4	1990	14,6	2005	13,8
1961	14,8	1976	13,8	1991	14,2	2006	14,7
1962	13,5	1977	14,1	1992	14,6	2007	15,2
1963	13,7	1978	13,5	1993	14,3	2008	14,5
1964	13,9	1979	13,9	1994	15,4	2009	14,9
1965	13,2	1980	13,3	1995	14,1	2010	13,7
1966	14,1	1981	13,8	1996	14,0	2011	15,1
1967	14,3	1982	14,2	1997	14,9	2012	15,0
1968	13,7	1983	14,0	1998	14,6	2013	14,7
1969	13,7	1984	13,3	1999	14,6	2014	14,7
1970	13,5	1985	13,7	2000	17,2	2015	15,6
1971	13,8	1986	13,9	2001	14,7		
1972	13,9	1987	13,6	2002	15,1		
1973	13,8	1988	14,4	2003	15,0		

(a) Temperature medie giornaliere



(b) Diagramma cartesiano

Figura 9 – Temperature medie annuali in gradi Celsius della stazione meteorologica di Venezia Cavanis. Anni 1959–2015. Dati forniti da Arpav. Fonte: http://seriestoriche.istat.it/fileadmin/documenti/Tavola_1.1.xls

```

1 import numpy as np
2 import matplotlib.pyplot as plt
3
4 # anni dal 1959 al 2015
5 x = np.arange(1959, 2016)
6 # elenco degli anni da visualizzare sull'asse x
7 xTicks = (1959, 1970, 1980, 1990, 2000, 2010, 2015)
8 # le 57 temperature medie rilevate
9 y = np.array([14.2, 13.7, 14.8, 13.5, 13.7, 13.9, 13.2, 14.1, 14.3, 13.7, 13.7, 13.5,
10              13.8, 13.9, 13.8, 13.9, 14.4, 13.8, 14.1, 13.5, 13.9, 13.3, 13.8, 14.2,
11              14.0, 13.3, 13.7, 13.9, 13.6, 14.4, 14.2, 14.6, 14.2, 14.6, 14.3, 15.4,
12              14.1, 14.0, 14.9, 14.6, 14.6, 17.2, 14.7, 15.1, 15.0, 14.2, 13.8, 14.7,
13              15.2, 14.5, 14.9, 13.7, 15.1, 15.0, 14.7, 14.7, 15.6])
14 fig, ax = plt.subplots(1, 1)
15 ax.plot(x, y)
16 ax.set_xticks(xTicks)
17 ax.set_ylabel('Temperatura media(°C)')
18 ax.set_title('Venezia Cavanis')
19 plt.show()

```

Programma 9 – Diagramma cartesiano

Riferimenti bibliografici

- [1] MASSIMO BERGAMINI, GRAZIELLA BAROZZI (2015). *Matematica multimediale.verde 1*, Zanichelli Editore, Bologna
- [2] C. ROSSI (1999). *La matematica dell'incertezza*, Zanichelli Editore, Bologna