

3

Rappresentazioni dei dati

3.1 Funzioni

Se esiste un concetto che caratterizza la matematica moderna è il concetto di funzione. Praticamente tutta la matematica (e le applicazioni della matematica non sono da meno) ne fa un uso continuo. Non ti sorprenderà quindi scoprire che dedicheremo un intero paragrafo a una discussione delle funzioni e delle loro proprietà.

Una *funzione* (o *applicazione*) fra due insiemi A e B è una legge che associa a ciascun elemento di A *uno e un solo* elemento di B . L'insieme di partenza A è il *dominio* della funzione; l'insieme di arrivo B il *codominio*. In simboli, una funzione f di dominio A e codominio B verrà indicata con $f: A \rightarrow B$. Se la funzione f manda l'elemento $a \in A$ nell'elemento $b \in B$, scriveremo¹ $b = f(a)$, e diremo che b è *immagine di a tramite f*. L'insieme degli elementi di B che sono immagine tramite f di elementi di A è l'*immagine di f*, e viene indicata con $\text{Im } f$ oppure con $f(A)$; in simboli,

$$f(A) = \{b \in B \mid b = f(a) \text{ per qualche } a \in A\} = \{f(a) \in B \mid a \in A\}.$$

Osservazione 3.1 Se $b = f(a)$ si dice che b è la *variabile dipendente*, mentre a è la *variabile indipendente*. Il significato di questa dizione è che mentre l'elemento a (che è anche chiamato *argomento* della funzione f) può essere un elemento qualsiasi del dominio, l'elemento $b = f(a)$ del codominio (che è anche chiamato *valore* della funzione nell'argomento a) è univocamente determinato dall'argomento a e dalla legge f ; in altre parole, il valore b *dipende* dall'argomento a della funzione, e il modo in cui ne dipende è esattamente la funzione f .

¹ Oppure $a \xrightarrow{f} b$, o semplicemente $a \mapsto b$, se il contesto individua chiaramente di quale funzione si tratta.

ESEMPIO 3.1 Il prezzo al litro della benzina senza piombo dal benzinaio all'angolo è una funzione che associa a ogni giorno degli ultimi tre anni un numero reale — ed è una funzione sfortunatamente non decrescente.

ESEMPIO 3.2 La legge che associa a ogni giorno dell'anno 2006 il prezzo al litro della benzina in Italia *non* è una funzione. Infatti, il prezzo cambia da benzinaio a benzinaio, per cui non si può associare a ciascun giorno dell'anno un unico prezzo.

ESEMPIO 3.3 La concentrazione di glucosio nel sangue di un paziente è una funzione, con dominio l'intervallo di tempo durante il quale viene effettuata la misura, e con codominio l'insieme dei numeri reali.

ESEMPIO 3.4 Il tuo peso è una funzione, con dominio l'intervallo di tempo dalla tua nascita alla tua morte, e con codominio l'insieme dei numeri reali positivi.

ESEMPIO 3.5 Una distribuzione di probabilità è una funzione, con dominio la famiglia \mathcal{A} di sottoinsiemi dello spazio degli eventi, e codominio l'intervallo $[0, 1]$.

ESEMPIO 3.6 La legge che associa a ciascuna persona il suo gruppo sanguigno è una funzione, con dominio l'insieme A di tutti gli esseri umani e codominio l'insieme $\{A, B, AB, 0\}$ dei possibili gruppi sanguigni.

ESEMPIO 3.7 La classifica della quinta giornata del girone d'andata del campionato di serie A del 2006/2007 è una funzione che associa a ogni squadra di serie A un numero naturale: il suo punteggio. Da questo punto di vista, il campionato consiste nel cambiare funzione ogni settimana.

ESEMPIO 3.8 Se A è un insieme qualsiasi, la funzione $\text{id}_A: A \rightarrow A$ che associa a ogni elemento di A se stesso (cioè $\text{id}_A(a) = a$ per ogni $a \in A$) è l'*identità* di A .

ESEMPIO 3.9 Siano A e B insiemi, e $b_0 \in B$ un elemento dato. La legge $f: A \rightarrow B$ che associa b_0 a ogni elemento di A (in simboli, $f(a) = b_0$ per ogni $a \in A$) è una funzione, detta *funzione costante* di valore b_0 .

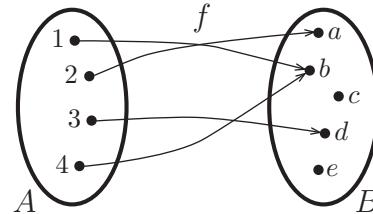


Figura 3.1 Rappresentazione grafica di una funzione.

ESEMPIO 3.10 Siano A e B gli insiemi $A = \{1, 2, 3, 4\}$ e $B = \{a, b, c, d, e\}$. La legge $f: A \rightarrow B$ data da $f(1) = b$, $f(2) = a$, $f(3) = d$, $f(4) = b$ è una fun-

zione. Possiamo rappresentarla con un disegno come in Figura 3.1. In questo caso, l'immagine $f(A) = \{a, b, d\}$ è strettamente più piccola del codominio.

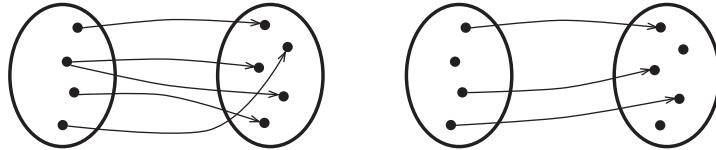


Figura 3.2 Leggi che non rappresentano funzioni.

Ogni funzione in cui sia il dominio che il codominio sono costituiti da un numero finito di punti può venire visualizzata con un disegno simile a quello della Figura 3.1: l'essenziale è che da *ogni* punto del dominio parta *una e una sola* freccia. Per intenderci, i due disegni della Figura 3.2 *non* rappresentano funzioni: il primo perché a un elemento del dominio vengono associati due elementi del codominio, e il secondo perché a un elemento del dominio non viene associato alcun elemento del codominio.

Osservazione 3.2 Possiamo pensare la funzione $f: A \rightarrow B$ come una specie di scatola nera, con un ingresso e un'uscita. Ogni volta che in ingresso entra un elemento del dominio, la scatola nera — la funzione — lo elabora e poi espelle dall'uscita un elemento del codominio. Non è importante la natura degli elementi del dominio e del codominio (possono essere numeri, rette, patate, cavalleggeri prussiani o qualsiasi altra cosa) né il tipo di processi digestivi che avvengono all'interno della scatola. Somme, prodotti, classifiche o formine da sabbia, tutto è ammissibile, *purché il procedimento usato sia sempre lo stesso*: ogni volta che in ingresso infiliamo la stessa patata, in uscita dobbiamo ottenere la stessa cipolla — a ogni elemento del dominio viene associato uno e un solo elemento del codominio, appunto.

Questa analogia ci permette di dire quando due funzioni sono uguali. Per i nostri scopi, due scatole nere che producono sempre lo stesso oggetto quando in ingresso ricevono lo stesso elemento sono indistinguibili: non potendo vedere come sono fatte dentro, se si comportano nello stesso modo per noi coincidono. Dunque diremo che due funzioni $f: A \rightarrow B$ e $g: A \rightarrow B$ (con lo stesso dominio e lo stesso codominio, s'intende) sono *uguali*, e scriveremo $f = g$ oppure $f \equiv g$, se e solo se $f(a) = g(a)$ per ogni $a \in A$.

ESEMPIO 3.11 Un tipo particolare di funzioni è costituito dai polinomi. Un *polinomio* (a coefficienti reali, in una variabile²) è una funzione $p: \mathbb{R} \rightarrow \mathbb{R}$ della forma

$$p(t) = a_n t^n + a_{n-1} t^{n-1} + \cdots + a_0, \quad (3.1)$$

² Una volta vista la definizione ti dovrebbe essere chiaro come costruire polinomi in più variabili.

dove $n \in \mathbb{N}$ è un numero naturale detto *grado* del polinomio, e $a_0, \dots, a_n \in \mathbb{R}$ sono numeri reali, i *coefficienti* del polinomio. In particolare, a_0 è detto *termine noto*, e $a_n \neq 0$ *coefficiente direttivo*. I polinomi di grado zero sono funzioni costanti. I polinomi di primo grado sono spesso detti *lineari*. Per esempio, $p(t) = 3t + 2$ è un polinomio lineare, mentre $q(t) = 6t^2 - 7t + 1$ è un polinomio di grado 2.

Il significato della scrittura (3.1) è il seguente: il polinomio p associa a ogni numero reale t il numero reale ottenuto prendendo t , elevandolo alla potenza n , moltiplicandolo per a_n , riprendendo t per moltiplicarlo per a_{n-1} dopo averlo elevato alla potenza $n-1$ in modo da poter sommare il risultato a quanto già avevamo, e così via, fino a sommare a_0 ottenendo finalmente il valore finale $p(t)$. Per esempio, se p è il polinomio dato da $p(t) = 2t^2 - t + 3\pi$, allora $p(-2) = 2(-2)^2 - (-2) + 3\pi = 10 + 3\pi$. Per noi, un polinomio sarà *sempre* una funzione, e *mai* una misteriosa combinazione formale di lettere e numeri. Indicheremo con $\mathbb{R}[t]$ l'insieme di tutti i polinomi (a coefficienti reali, in una variabile), e con $\mathbb{R}_n[t]$ l'insieme dei polinomi di grado minore o uguale a $n \in \mathbb{N}$. Torneremo a parlare dei polinomi nel prossimo capitolo.

ESEMPIO 3.12 Una parte consistente di questo corso sarà dedicato allo studio delle *funzioni reali di una variabile reale*, cioè alle funzioni f con dominio A contenuto in \mathbb{R} (avremo casi in cui $A = \mathbb{R}$ e casi in cui A sarà strettamente più piccolo di \mathbb{R}), e a valori in \mathbb{R} (cioè con codominio \mathbb{R}). I polinomi, per esempio, sono funzioni reali di variabile reale; e lo sono anche le funzioni razionali, le funzioni trigonometriche seno e coseno, come pure l'esponenziale e il logaritmo³. Nota però che, mentre polinomi, seno, coseno ed esponenziale sono definiti su tutto l'asse reale (cioè hanno dominio uguale a \mathbb{R}), le funzioni razionali e il logaritmo in generale non lo sono. Per la precisione, il dominio di una funzione razionale non può contenere gli zeri del denominatore, e il logaritmo ha come dominio l'insieme dei numeri reali positivi.

ESEMPIO 3.13 Vi sono ovviamente moltissime altre funzioni da \mathbb{R} in \mathbb{R} , di cui alcune discretamente orrende. Un esempio di questo genere è la *scala del diavolo*, cioè la funzione $f: \mathbb{R} \rightarrow \mathbb{R}$ definita da

$$f(x) = \begin{cases} x & \text{se } x \text{ è razionale,} \\ 0 & \text{se } x \text{ è irrazionale.} \end{cases}$$

Torniamo a parlare di concetti generali. Sia $f: A \rightarrow B$ una funzione. Se $A_1 \subseteq A$ è un sottoinsieme di A , l'*immagine di A_1 tramite f* è l'insieme $f(A_1) \subseteq B$ delle immagini degli elementi di A_1 . Viceversa, se $B_1 \subseteq B$, l'insieme degli elementi di A la cui immagine tramite f appartiene a B_1 si chiama *immagine inversa* $f^{-1}(B_1)$ di B_1 tramite f ; in simboli,

$$f^{-1}(B_1) = \{a \in A \mid f(a) \in B_1\} \subseteq A.$$

Chiaramente, $f^{-1}(B) = A$ per qualunque funzione $f: A \rightarrow B$ (perché?).

³ Parleremo in dettaglio di tutte queste funzioni nel prossimo capitolo.

ESEMPIO 3.14 Sia $f: A \rightarrow B$ la funzione definita nell’Esempio 3.10, e prendiamo $A_1 = \{2, 4\}$ e $B_1 = \{b, d, e\}$. Allora $f(A_1) = \{a, b\}$ e $f^{-1}(B_1) = \{1, 3, 4\}$.

ESEMPIO 3.15 Sia f la funzione che associa a ogni distributore di benzina d’Italia il prezzo al litro della benzina senza piombo il 4 maggio 2006. Se A_1 è il sottoinsieme dei distributori di benzina in Toscana, allora $f(A_1)$ è il sottoinsieme dei prezzi della benzina senza piombo in Toscana il 4 maggio 2006. Se $B_1 \subset \mathbb{R}$ è il sottoinsieme dei numeri reali minori di 1.1, allora $f^{-1}(B_1)$ è l’insieme dei distributori d’Italia in cui il 4 maggio 2006 la benzina senza piombo costava meno di 1.1 euro al litro.

Osservazione 3.3 Una funzione con dominio l’insieme dei numeri naturali è spesso chiamata *successione*. Infatti dare una funzione $a: \mathbb{N} \rightarrow A$ significa scegliere una sequenza di elementi del codominio, cominciando da $a(0)$ e poi proseguendo con $a(1)$, $a(2)$ e così via. Di solito in questo caso si scrive a_n invece di $a(n)$. Ripareremo di successioni in un prossimo capitolo.

Osservazione 3.4 Stai attento a non confondere il concetto di funzione con quello di equazione. Come detto, una funzione $f: A \rightarrow B$ è un modo per associare a ogni elemento del dominio A un elemento del codominio B . Invece, un’equazione è una *domanda* associata alla f e a uno specifico elemento $b \in B$ del codominio. Infatti, studiare l’equazione $f(x) = b$ vuol dire chiedersi se esistono elementi $a \in A$ del dominio che abbiano immagine b , cioè tali che $f(a) = b$. In altre parole, risolvere l’equazione $f(x) = b$ equivale a trovare l’immagine inversa $f^{-1}(\{b\})$ tramite f del sottoinsieme $\{b\}$; gli elementi di $f^{-1}(\{b\})$ sono le *soluzioni* dell’equazione $f(x) = b$. Ovviamente, se b non appartiene all’immagine di f allora $f^{-1}(\{b\}) = \emptyset$, cioè l’equazione $f(x) = b$ non ha soluzioni.

ESEMPIO 3.16 Sia $p: \mathbb{R} \rightarrow \mathbb{R}$ il polinomio $p(t) = t^2 - 1$. Se $c \in \mathbb{R}$ è un qualsiasi numero reale, risolvere l’equazione $p(x) = c$ equivale a trovare (se esistono) dei numeri reali che diano risultato c elevandoli al quadrato e sottraendo 1. Per esempio, l’equazione $x^2 - 1 = 0$ ha come soluzione i numeri 1 e -1, cioè $p^{-1}(\{0\}) = \{1, -1\}$. Invece, $p^{-1}(\{-2\}) = \emptyset$, cioè l’equazione $x^2 - 1 = -2$ non ha soluzioni (in quanto $x^2 - 1$ è sempre maggiore o uguale di -1). Infine, l’equazione $x^2 - 1 = -1$ ha come unica soluzione $x = 0$, cioè $p^{-1}(\{-1\}) = \{0\}$.

Osservazione 3.5 Sia $f: A \rightarrow B$ una funzione da A a B . Se A_1 è un sottoinsieme di A , la funzione f chiaramente determina anche una legge che a ogni elemento di A_1 associa un elemento di B (la stessa legge di prima), e quindi una funzione da A_1 a B . Questa nuova funzione si chiama *restrizione* di f ad A_1 , e si indica con $f|_{A_1}$. Se capiterà, scriveremo $f(A_1)$ e non $f|_{A_1}(A_1)$ per indicare l’immagine di f ristretta ad A_1 , in modo da non complicare troppo le formule.

ESEMPIO 3.17 La funzione $f = p|_{\mathbb{N}}: \mathbb{N} \rightarrow \mathbb{R}$ che associa a ogni numero naturale il suo successore è la restrizione a \mathbb{N} del polinomio $p: \mathbb{R} \rightarrow \mathbb{R}$ dato da $p(t) = t + 1$.

ESEMPIO 3.18 La funzione che ci fornisce la concentrazione del glucosio nel sangue di un paziente ogni ora è una restrizione della funzione dell’Esempio 3.3 che ci

fornisce la concentrazione del glucosio a ogni istante.

Osservazione 3.6 Cambiando il dominio di una funzione cambiano le soluzioni delle equazioni che possiamo associare alla funzione. Per esempio, sia $p: \mathbb{R} \rightarrow \mathbb{R}$ il polinomio $p(x) = x^2 - 1$ visto nell’Esempio 3.16, e indichiamo con $q = p|_{\mathbb{R}^+}: \mathbb{R}^+ \rightarrow \mathbb{R}$ la sua restrizione ai soli numeri reali positivi. Allora l’equazione $q(x) = 0$ ha come unica soluzione $x = 1$, in quanto $q^{-1}(\{0\})$ dev’essere un sottoinsieme del dominio di q , cioè \mathbb{R}^+ .

Nel definire il concetto di funzione, abbiamo sottolineato che è importante che a *ogni* elemento del dominio venga associato *esattamente un* elemento del codominio; se questo non avviene non si può parlare di funzione. Richiedendo proprietà analoghe per gli elementi del codominio possiamo invece identificare alcune classi particolarmente significative di funzioni.

Per esempio, non sempre ogni elemento del codominio è associato a (è immagine di) un elemento del dominio; in generale, l’immagine è un sottoinsieme proprio del codominio. Se invece $f: A \rightarrow B$ è una funzione tale che ogni elemento del codominio è immagine di (almeno) un elemento del dominio, cioè $\text{Im } f = B$, diremo che f è *surgettiva*.



Figura 3.3 Funzioni surgettive e non surgettive.

In termini della rappresentazione grafica introdotta nell’Esempio 3.10, una funzione è surgettiva se ogni elemento del codominio è raggiunto da almeno una freccia: nella Figura 3.3 la funzione a sinistra è surgettiva, quella a destra no.

Osservazione 3.7 In termini di equazioni, dire che una funzione $f: A \rightarrow B$ è surgettiva equivale (perché?) a dire che per qualsiasi $b \in B$ l’equazione $f(x) = b$ ha almeno una soluzione.

ESEMPIO 3.19 Il polinomio $p: \mathbb{R} \rightarrow \mathbb{R}$ dato da $p(t) = 2t - 1$ è surgettivo. Infatti, per ogni $b \in \mathbb{R}$ si ha $f((b+1)/2) = b$, cioè l’elemento b del codominio è immagine dell’elemento $(b+1)/2 \in \mathbb{R}$ del dominio. Invece, il polinomio $q: \mathbb{R} \rightarrow \mathbb{R}$ dato da $q(t) = t^2 + 1$ non è surgettivo, in quanto l’equazione $q(x) = 0$ non ha soluzioni (cioè 0 non è nell’immagine di q).

Un’altra cosa che può accadere è che ci siano elementi del codominio che siano associati a più di un elemento del dominio, cioè che due elementi diversi del dominio abbiano la stessa immagine. Se invece $f: A \rightarrow B$ è tale che ogni elemento del codominio è immagine di *al più* un elemento del dominio, cioè se f associa elementi

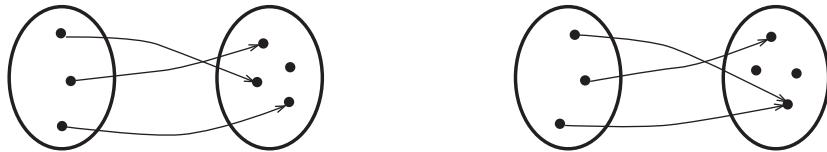


Figura 3.4 Funzioni iniettive e non iniettive.

diversi del codominio a elementi diversi del dominio — cioè se, in simboli, $a_1 \neq a_2$ implica $f(a_1) \neq f(a_2)$ — diremo che f è *iniettiva*.

In termini della solita rappresentazione grafica, una funzione è iniettiva se su ogni elemento del codominio arriva al più una freccia (ma può anche non arrivarne alcuna): nella Figura 3.4 la funzione a sinistra è iniettiva e quella a destra no.

Osservazione 3.8 In termini di equazioni, dire che una funzione $f: A \rightarrow B$ è iniettiva equivale (perché?) a dire che per qualsiasi $b \in B$ l'equazione $f(x) = b$ ha al massimo una soluzione (ma può anche non averne).

ESEMPIO 3.20 Il polinomio $p: \mathbb{R} \rightarrow \mathbb{R}$ dato da $p(t) = 2t - 1$ è iniettivo. Infatti, se $a_1, a_2 \in \mathbb{R}$ sono tali che $p(a_1) = p(a_2)$ si deve avere $2a_1 - 1 = 2a_2 - 1$, cioè $2a_1 = 2a_2$, e quindi $a_1 = a_2$. Invece, il polinomio $q: \mathbb{R} \rightarrow \mathbb{R}$ dato da $q(t) = t^2 + 1$ non è iniettivo, in quanto l'equazione $q(x) = 2$ ha due soluzioni 1 e -1, cioè $q(1) = 2 = q(-1)$.

Una funzione $f: A \rightarrow B$ iniettiva e surgettiva verrà detta *bigettiva* (o *bijettiva*, o *biunivoca*). Tramite una funzione bigettiva, ciascun elemento del codominio è associato a uno e un solo elemento del dominio⁴; se $b \in B$, esiste un *unico* $a \in A$ tale che $f(a) = b$. Questo ci permette di definire una funzione da B ad A , la *funzione inversa* $f^{-1}: B \rightarrow A$, ponendo $f^{-1}(b) = a$, dove $a \in A$ è quell'unico elemento tale che $f(a) = b$. Le funzioni bigettive si dicono anche *invertibili*.

Usando di nuovo la rappresentazione dell'Esempio 3.10, una funzione è bigettiva se su ogni elemento del codominio arriva esattamente una freccia, e in tal caso la funzione inversa si ottiene invertendo il senso delle frecce. Per intenderci, nessuna delle funzioni nelle Figure 3.3 e 3.4 era bigettiva, mentre la Figura 3.5 ci mostra una funzione invertibile (a sinistra) assieme alla sua inversa (a destra).

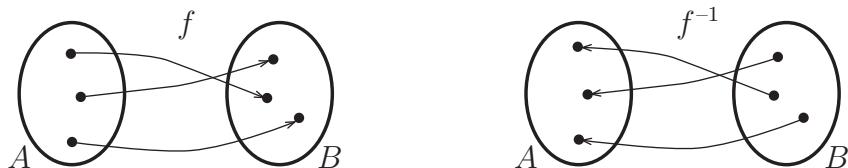


Figura 3.5 Una funzione bigettiva e la sua inversa.

⁴ In un certo senso, questo vuol dire che il dominio e il codominio hanno lo stesso numero di elementi; in A ce ne sono tanti quanti in B . Si dice anche che f è una *corrispondenza biunivoca* fra A e B .

Osservazione 3.9 In termini di equazioni, dire che una funzione $f: A \rightarrow B$ è invertibile equivale (perché?) a dire che per qualsiasi $b \in B$ l'equazione $f(x) = b$ ha esattamente una soluzione. La funzione inversa è ottenuta associando a ciascun $b \in B$ l'unica soluzione $a \in A$ dell'equazione $f(x) = b$.

ESEMPIO 3.21 Dagli esempi precedenti deduciamo che il polinomio $p: \mathbb{R} \rightarrow \mathbb{R}$ dato da $p(t) = 2t - 1$ è bigettivo, e la funzione inversa $p^{-1}: \mathbb{R} \rightarrow \mathbb{R}$ è data da $p^{-1}(t) = (t + 1)/2$.

Osservazione 3.10 Attenzione: l'essere o meno surgettiva (o iniettiva) dipende fortemente dal dominio e dal codominio di una funzione. Per esempio, se $f: A \rightarrow B$ è una funzione non surgettiva, la funzione $f_1: A \rightarrow \text{Im } f$ data dalla stessa legge di f (cioè $f_1(a) = f(a)$ per ogni $a \in A$) è surgettiva, in quanto abbiamo tolto dal codominio tutti gli elementi di B che non stavano nell'immagine. Se invece $g: A \rightarrow B$ è una funzione non iniettiva, scegliamo per ogni elemento $b \in \text{Im } g$ un specifico elemento del dominio A con immagine b , e indichiamo con A_1 il sottoinsieme di A composto dagli elementi così scelti. Allora $g|_{A_1}: A_1 \rightarrow B$ è iniettiva, in quanto abbiamo tolto dal dominio tutti gli elementi che hanno la stessa immagine di elementi di A_1 .

ESEMPIO 3.22 Abbiamo visto che il polinomio $q: \mathbb{R} \rightarrow \mathbb{R}$ dato da $q(t) = t^2 + 1$ non è né iniettivo né surgettivo. Si vede subito (perché?) che l'immagine di q è la semiretta $[1, +\infty)$; quindi la funzione $q_1: \mathbb{R} \rightarrow [1, +\infty)$ data da $q_1(t) = t^2 + 1$ è surgettiva. Inoltre, ogni elemento di $[1, +\infty)$ è immagine tramite q_1 di un unico numero non negativo; quindi la restrizione $q_2 = q_1|_{[0, +\infty)}: [0, +\infty) \rightarrow [1, +\infty)$ è iniettiva e surgettiva, cioè bigettiva.

Osservazione 3.11 Attenzione a non confondere i concetti di funzione inversa e di immagine inversa. La funzione inversa f^{-1} associa a ogni elemento del codominio di f un elemento del dominio di f , ed esiste soltanto quando la funzione f è bigettiva. L'immagine inversa, invece, associa a un sottoinsieme del codominio un sottoinsieme del dominio — per cui *non* è una funzione definita sul codominio — ed esiste sempre, anche quando la funzione f non è bigettiva.

Osservazione 3.12 Attenzione anche a non confondere i concetti di funzione e di funzione iniettiva. Una funzione $f: A \rightarrow B$ associa sempre *a ogni elemento di A uno e un solo elemento di B*; per una funzione iniettiva invece *ogni elemento di B è immagine di al più un elemento di A*, che è un concetto ben diverso. In una funzione qualunque, da ogni elemento del dominio *parte* esattamente una freccia; in una funzione iniettiva, su ogni elemento del codominio *arriva* al più una freccia.

Supponiamo ora di avere due funzioni $f: A \rightarrow B$ e $g: B \rightarrow C$, dove il codominio di f coincide col dominio di g . In tal caso possiamo definire una nuova funzione, la *composizione* $g \circ f: A \rightarrow C$ delle funzioni f e g , tramite la formula

$$\forall a \in A \quad (g \circ f)(a) = g(f(a)) . \quad (3.2)$$

La $g \circ f$ è effettivamente una funzione: infatti, f associa a ciascun $a \in A$ un unico elemento di B , e g associa a quest'ultimo un unico elemento di C , per cui $g \circ f$ è una legge che a ciascun elemento di A associa uno e un solo elemento di C .

Osservazione 3.13 Attenzione: nella composizione $g \circ f$ *prima* si applica f e *poi* si applica g . Infatti, per calcolare $g \circ f(a) = g(f(a))$ dobbiamo prima trovare $f(a)$ e poi l'immagine di $f(a)$ tramite g .

Osservazione 3.14 Più in generale, la composizione $g \circ f$ di due funzioni $f: A \rightarrow B$ e $g: B \rightarrow C$ è ben definita non appena l'immagine di f è contenuta nel dominio di g . Infatti in tal caso il secondo membro di (3.2) è ben definito per ogni $a \in A$.

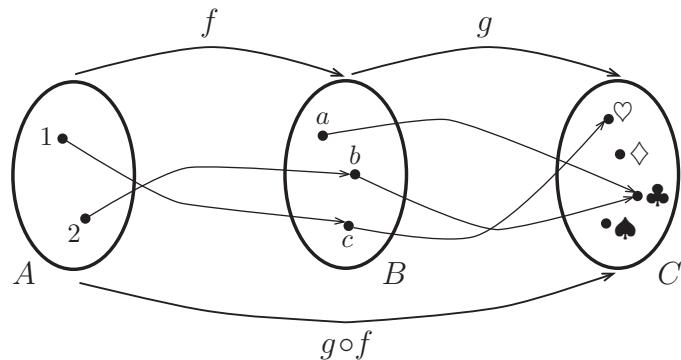


Figura 3.6 Composizione di funzioni.

ESEMPIO 3.23 Prendiamo i tre insiemi $A = \{1, 2\}$, $B = \{a, b, c\}$, $C = \{\heartsuit, \diamondsuit, \clubsuit, \spadesuit\}$ e le funzioni $f: A \rightarrow B$ e $g: B \rightarrow C$ date da $f(1) = c$, $f(2) = b$ e da $g(a) = \clubsuit = g(b)$, e $g(c) = \heartsuit$. Allora la composizione di f e g è la funzione $g \circ f: A \rightarrow C$ data da $(g \circ f)(1) = \heartsuit$ e $(g \circ f)(2) = \clubsuit$ (vedi la Figura 3.6).

ESEMPIO 3.24 Siano $f: \mathbb{N} \rightarrow \mathbb{R}$ e $g: \mathbb{R} \rightarrow \mathbb{R}$ le funzioni date da $f(n) = 2n + 3$ e da $g(x) = x^2$. Allora la composizione di f e g è la funzione $g \circ f: \mathbb{N} \rightarrow \mathbb{R}$ data da $(g \circ f)(n) = g(f(n)) = g(2n + 3) = (2n + 3)^2$. Nota che l'immagine di $g \circ f$ è contenuta in \mathbb{N} , anche se l'immagine di g è ben più grande (è l'insieme di tutti i numeri reali non negativi).

Osservazione 3.15 Attenzione: la composizione non è commutativa. Se $g \circ f$ è definita, non è detto che $f \circ g$ lo sia, e anche se lo fosse potrebbe essere alquanto diversa da $g \circ f$.

ESEMPIO 3.25 Sia f la funzione che associa a ogni città d'Italia il suo sindaco, e g la funzione che associa a ogni essere umano il suo gruppo sanguigno. Allora $g \circ f$ è la funzione che associa a ogni città il gruppo sanguigno del suo sindaco, ma $f \circ g$ si guarda bene dall'essere definita. Infatti, calcolare $f \circ g$ richiederebbe di trovare il sindaco di un gruppo sanguigno...

ESEMPIO 3.26 Siano $p, q: \mathbb{R} \rightarrow \mathbb{R}$ i polinomi $p(t) = 2t - 1$ e $q(t) = t^2 + 1$. Allora

$$q(p(t)) = q(2t - 1) = (2t - 1)^2 + 1 = 4t^2 - 4t + 2$$

mentre

$$p(q(t)) = p(t^2 + 1) = 2(t^2 + 1) - 1 = 2t^2 + 1,$$

per cui $q \circ p$ e $p \circ q$, pur essendo entrambi definiti, sono polinomi alquanto diversi.

Esercizio 3.1 Dati $a, b \in \mathbb{R}$, indichiamo con $p_{a,b}: \mathbb{R} \rightarrow \mathbb{R}$ il polinomio lineare $p_{a,b}(t) = at + b$. Per quali valori di a e b il polinomio $p_{a,b}$ è iniettivo? Per quali surgettivo? Per quali bigettivo?

Esercizio 3.2 Siano $f: A \rightarrow B$ e $g: B \rightarrow C$ due funzioni tali che $g \circ f: A \rightarrow C$ sia iniettiva. Puoi dedurre che f è iniettiva? O che g è iniettiva? Perché?

A ogni funzione $f: A \rightarrow B$ possiamo associare un sottoinsieme del prodotto cartesiano del dominio col codominio che rappresenta bene la funzione, e da cui (come vedremo) saremo in grado di leggere proprietà della funzione stessa.

Sia allora $f: A \rightarrow B$ una funzione, di dominio A e codominio B . Il grafico Γ_f di f è il sottoinsieme del prodotto cartesiano $A \times B$ dato da

$$\Gamma_f = \{(a, b) \in A \times B \mid b = f(a)\} = \{(a, f(a)) \mid a \in A\} \subseteq A \times B.$$

ESEMPIO 3.27 Il grafico della funzione $f: A \rightarrow B$ dell'Esempio 3.10 è dato da

$$\Gamma_f = \{(1, b), (2, a), (3, d), (4, b)\}.$$

Infatti, $f(1) = b$, $f(2) = a$, $f(3) = d$ e $f(4) = b$.

ESEMPIO 3.28 Il grafico della concentrazione di glucosio nel sangue di un paziente è l'insieme delle coppie ordinate $(t, c) \in \mathbb{R}^2$ di numeri reali per cui c è la concentrazione di glucosio misurata al tempo t .

ESEMPIO 3.29 Il grafico della funzione gruppo sanguigno è l'insieme delle coppie ordinate (p, G) tali che G sia il gruppo sanguigno della persona p .

Non ogni sottoinsieme di un prodotto cartesiano $A \times B$ può essere il grafico di una funzione da A a B . Infatti, un sottoinsieme $\Gamma \subseteq A \times B$ è il grafico di una funzione f se e solo se (perché?) per ogni $a \in A$ esiste un unico elemento $b \in B$ tale che (a, b) in Γ , e in tal caso $b = f(a)$.

Esercizio 3.3 Sia $\Gamma \subseteq A \times B$ un sottoinsieme (non vuoto) di un prodotto cartesiano tale che per ogni $a \in A$ esiste al più un elemento $b \in B$ (ma potrebbe non esisterne alcuno) tale che $(a, b) \in \Gamma$. Indica con $A_1 \subseteq A$ l'insieme degli elementi $a \in A$ per cui esiste un $b \in B$ con $(a, b) \in \Gamma$. Mostra che $\Gamma \subseteq A_1 \times B$, e che Γ è il grafico di una funzione da A_1 a B .

Se mettiamo gli elementi di A su un'ipotetica linea orizzontale, e quelli di B su una ipotetica linea verticale, allora gli elementi di $A \times B$ corrispondono a tutte le intersezioni delle linee verticali passanti per elementi di A con le linee orizzontali passanti per elementi di B . In questa rappresentazione grafica, un sottoinsieme $\Gamma \subseteq A \times B$ è un grafico se e solo se ogni linea verticale passante per un elemento di A interseca Γ in uno e un solo punto (perché?).

Considerando invece le intersezioni delle linee orizzontali con il grafico possiamo identificare alcune proprietà della funzione. Per esempio, una funzione $f: A \rightarrow B$ è surgettiva se e solo se per ogni $b \in B$ esiste almeno un elemento $a \in A$ tale che $b = f(a)$, o, in altre parole, tale che $(a, b) \in \Gamma_f$. Quindi (perché?) f è surgettiva se e solo se ogni linea orizzontale passante per un elemento di B interseca il grafico di f . Più in generale, l'immagine di f consiste negli elementi $b \in B$ tali che la linea orizzontale passante per b interseca il grafico di f .

Analogamente, una funzione $f: A \rightarrow B$ è iniettiva se e solo se per ogni $b \in B$ esiste al massimo un elemento $a \in A$ tale che $b = f(a)$, o, in altre parole, tale che $(a, b) \in \Gamma_f$. Quindi (perché?) f è iniettiva se e solo se ogni linea orizzontale passante per un elemento di B interseca il grafico di f in al massimo un punto.

Osservazione 3.16 Più in generale, diremo che un sottoinsieme qualsiasi $R \subseteq A \times B$ rappresenta una *relazione* fra gli elementi di A e gli elementi di B , nel senso che $a \in A$ è in relazione con $b \in B$ se e solo se $(a, b) \in R$.

ESEMPIO 3.30 Se A è l'insieme dei neuroni del tuo cervello, il sottoinsieme $R \subset A \times A$ delle coppie (a_1, a_2) di neuroni tali che il neurone a_1 sia in grado di mandare un segnale elettrico al neurone a_2 rappresenta una relazione sull'insieme dei neuroni del tuo cervello.

ESEMPIO 3.31 Se A è l'insieme degli esseri umani vivi il 9 dicembre 2006, allora la relazione “essere sposato a” è rappresentata dal sottoinsieme $R \subset A \times A$ costituito dalle coppie (ordinate) di coniugi.

ESEMPIO 3.32 Sia A l'insieme degli elementi chimici, e $B = \mathbb{N}$. Allora possiamo considerare la relazione $R \subset A \times \mathbb{N}$ costituita dalle coppie $(a, n) \in A \times \mathbb{N}$, dove n è il peso atomico (arrotondato all'intero più vicino) di un isotopo dell'elemento a . Siccome ogni elemento ha isotopi con pesi atomici diversi (per esempio, l'idrogeno ha isotopi con peso atomico 1, 2 e 3), questa relazione non è il grafico di una funzione.

ESEMPIO 3.33 L'insieme delle coppie (x, y) di numeri reali tali che $x^2 + y^2 > 1$ rappresenta una relazione sull'insieme di numeri reali.

3.2 Coordinate cartesiane

Nel seguito di questo corso ci occuperemo spesso di funzioni con dominio e codominio contenuti nella retta reale \mathbb{R} (funzioni reali di variabile reale). Il grafico di

queste funzioni è contenuto in $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$; obiettivo di questa sezione è richiamare le idee di base della geometria analitica, che permettono di identificare \mathbb{R}^2 con il piano euclideo e quindi rappresentare i grafici di funzioni reali di variabile reale come sottoinsiemi del piano.

Nella Sezione 1.1 abbiamo visto come costruire una corrispondenza biunivoca fra i punti di una retta e i numeri reali; vediamo ora come usare una procedura analoga per costruire una corrispondenza biunivoca fra i punti del piano e le coppie ordinate di numeri reali.

Il primo passo consiste anche stavolta nella scelta di un punto O del piano che chiameremo *origine*; sarà il punto che verrà associato alla coppia $(0, 0)$, e che quindi sarà in un certo senso il centro della nostra rappresentazione.

Il secondo passo consiste nella scelta di una retta r_1 qualsiasi passante per l'origine; sarà l'*asse delle ascisse*.

Il terzo passo consiste nella scelta di un punto A_1 distinto dall'origine sull'asse delle ascisse. Come abbiamo visto nella Sezione 1.1, questa scelta fornisce un'unità di misura e un'orientazione sull'asse delle ascisse r_1 , e quindi una corrispondenza biunivoca fra i punti di r_1 e i numeri reali. Nota che la retta r_1 è l'unica retta passante per O e A_1 ; quindi l'asse delle ascisse è completamente determinato da (l'origine e dal) punto A_1 .

Il quarto passo consiste nella scelta dell'*asse delle ordinate*, una seconda retta r_2 passante per l'origine e distinta da r_1 . Usualmente si sceglie l'asse delle ordinate ortogonale all'asse delle ascisse, ma non è strettamente necessario; la costruzione che stiamo descrivendo funziona con qualsiasi retta per l'origine diversa da r_1 .

Il quinto e ultimo passo consiste nella scelta di un punto A_2 distinto dall'origine sull'asse delle ordinate r_2 , ottenendo un'unità di misura e un'orientazione anche su r_2 (e l'asse delle ordinate è completamente determinato dalla scelta del punto A_2 fuori dall'asse delle ascisse).

Come vedremo fra un secondo, la scelta dei tre punti O , A_1 e A_2 , o, come diremo, la scelta del *sistema di riferimento* (o *di coordinate*) $R(O, A_1, A_2)$, determina una corrispondenza biunivoca fra punti del piano e coppie ordinate di numeri reali.

Osservazione 3.17 Se l'unità di misura sull'asse delle ascisse (la lunghezza del segmento $\overline{OA_1}$) è diversa da quella sull'asse delle ordinate (la lunghezza del segmento $\overline{OA_2}$), si parla di sistema di riferimento *dimetrico*; se invece sono uguali si parla di sistema di riferimento *monometrico*. La scelta di un sistema di riferimento monometrico o di uno dimetrico dipende molto dal tipo di problema che si deve affrontare; vedi l'Osservazione 3.20.

Scegliamo ora un punto P del piano; dobbiamo associargli una coppia di numeri reali. La procedura è la seguente: cominciamo tracciando la retta passante per P e parallela all'asse delle ordinate. Questa retta interseca l'asse delle ascisse in un unico punto P_1 , a cui corrisponde (grazie alla scelta dell'unità di misura) un unico numero reale $x \in \mathbb{R}$, l'*ascissa* del punto P .

Analogamente, la retta passante per P e parallela all'asse delle ordinate interseca l'asse delle ordinate in un unico punto P_2 , a cui corrisponde un unico numero reale $y \in \mathbb{R}$, l'*ordinata* del punto P . La coppia $(x, y) \in \mathbb{R}^2$ sono le *coordinate*

cartesiane del punto P rispetto al sistema di riferimento $R(O, A_1, A_2)$; vedi la Figura 3.7.

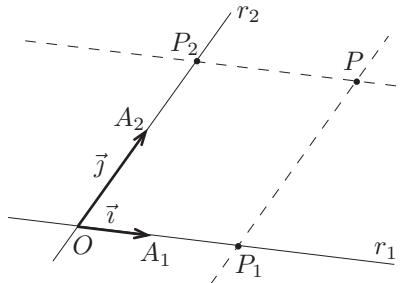


Figura 3.7 Coordinate nel piano.

Abbiamo quindi costruito una funzione $P \xrightarrow{f} (x, y)$ che associa a ogni punto del piano una ben definita coppia di numeri reali. Per verificare che è una corrispondenza biunivoca, ci basta costruire la funzione inversa, cioè una funzione g che associa a ogni coppia (x, y) di numeri reali un punto del piano in modo che sia $g \circ f$ sia $f \circ g$ siano l'identità.

Per costruire g ci basta invertire la precedente procedura. Data la coppia di numeri (x, y) , sia $P_1 \in r_1$ il punto dell'asse delle ascisse associato a x , e $P_2 \in r_2$ il punto dell'asse delle ordinate associato a y . Allora la retta passante per P_1 parallela all'asse delle ordinate interseca la retta passante per P_2 parallela all'asse delle ascisse in un unico punto P ; poniamo $g(x, y) = P$. In questo modo abbiamo ottenuto una funzione g che associa a ogni coppia di numeri reali un punto del piano. Non è difficile vedere (controlla) che g è effettivamente l'inversa di f , per cui abbiamo trovato una corrispondenza biunivoca fra i punti del piano e le coppie di numeri reali.

Osservazione 3.18 Nello spazio, scegliendo un terzo asse r_3 , non contenuto nel piano determinato dalle due rette r_1 ed r_2 , e un punto $A_3 \in r_3$ si ottiene (esercizio, non troppo difficile, per te) una corrispondenza biunivoca fra i punti dello spazio e le terne ordinate di numeri reali.

Osservazione 3.19 Quale sistema di riferimento scegliere dipende dal problema che si deve affrontare. Se è un problema di tipo geometrico avente a che fare con figure nel piano, allora potrebbe essere il problema stesso a suggerire le scelte migliori. Per esempio, se si deve studiare una circonferenza di solito conviene scegliere il centro della circonferenza come origine del sistema di riferimento; se si deve studiare un parallelogramma conviene scegliere gli assi paralleli ai lati del parallelogramma. Nel seguito, però, noi ci troveremo molto spesso a seguire la strada opposta: partiremo da un insieme di coppie di numeri reali (per esempio, il grafico di una funzione reale di variabile reale) e vogliamo associarigli un sottoinsieme del piano. In tal caso, la scelta del sistema di riferimento dev'essere fatta in modo che l'insieme (il grafico) sia rappresentato nella maniera più fedele possibile. Per esempio, tutto l'insieme (o

almeno la sua parte più significativa) dev'essere contenuto nella porzione di piano a noi visibile, occupandola in buona parte. Per intenderci, se ascissa e ordinata delle coppie di numeri che vogliamo rappresentare nel piano variano fra 10 e 20, e abbiamo a disposizione un intero foglio formato A4, potremo mettere l'origine in basso a sinistra nel foglio, scegliere assi ortogonali e unità di misura di circa un centimetro. Ma se ascisse e ordinate variassero fra 10^5 e $2 \cdot 10^5$, questa scelta rappresenterebbe l'insieme a un paio di chilometri dal nostro foglio; e se ascisse e ordinate variassero fra 10^{-1} e $2 \cdot 10^{-1}$ con questa scelta il nostro insieme verrebbe rappresentato in un quadratino di un millimetro di lato.

Osservazione 3.20 Anche la scelta fra sistemi di riferimento monometrici e dimetrici dipende da cosa vogliamo rappresentare. Tipicamente, se ascisse e ordinate si riferiscono a quantità dello stesso tipo e dello stesso ordine di grandezza i sistemi monometrici sono più convenienti; ma se ascisse e ordinate si riferiscono a grandezze di tipo diverso (per esempio, tempo e concentrazione del glucosio nel sangue) allora un sistema dimetrico può essere più efficiente. *Attenzione:* passare da un sistema monometrico a un sistema dimetrico cambia l'aspetto degli insiemi raffigurati. Per esempio, un quadrato in un sistema monometrico diventa un rettangolo in un sistema dimetrico, una circonferenza diventa un'ellisse, le rette cambiano pendenza, e così via. Anche l'aspetto di un grafico cambia passando da un sistema di riferimento a un altro (per quanto le informazioni contenute siano sempre le stesse); è importante che tu scelga il sistema di riferimento in modo che il disegno evidensi al meglio le caratteristiche che tu ritieni più importanti in quel grafico.

CURIOSITÀ 3.1 Scegliendo un sistema di riferimento diverso, allo stesso punto P vengono associate coordinate diverse; che relazione c'è fra loro? Indichiamo con \mathcal{A}^2 il piano euclideo, e con $\mathcal{R} = R(O, A_1, A_2)$ e $\mathcal{R}' = R(O', A'_1, A'_2)$ due sistemi di riferimento sul piano. Il sistema \mathcal{R} ci fornisce una bigezione $f: \mathcal{A}^2 \rightarrow \mathbb{R}^2$, che associa a ciascun punto $P \in \mathcal{A}^2$ la coppia di coordinate $f(P) = (x, y)$. Analogamente, il sistema \mathcal{R}' ci fornisce una seconda bigezione $f': \mathcal{A}^2 \rightarrow \mathbb{R}^2$, che associa al punto $P \in \mathcal{A}^2$ le coordinate $f'(P) = (x', y')$. La composizione $f' \circ f^{-1}: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ è quindi una bigezione di \mathbb{R}^2 con se stesso, che manda le coordinate (x, y) nelle coordinate (x', y') ; in altre parole, la funzione $f' \circ f^{-1}$ è il *cambiamento di coordinate* dal sistema \mathcal{R} al sistema \mathcal{R}' . Si può dimostrare che

$$\begin{cases} x' = b_{11}x + b_{12}y + c_1, \\ y' = b_{21}x + b_{22}y + c_2, \end{cases}$$

dove $(c_1, c_2) = f'(O)$ sono le coordinate di O nel sistema di riferimento \mathcal{R}' , e $b_{11}, b_{12}, b_{21}, b_{22}$ sono opportuni numeri reali.

Osservazione 3.21 Dunque una volta fissato un sistema di riferimento abbiamo una corrispondenza biunivoca fra punti del piano e coppie ordinate di numeri reali. Per questo motivo spesso, con un lieve abuso di terminologia, useremo frasi del tipo “il punto $P = (x, y)$ ” invece di “il punto P di coordinate (x, y) rispetto al sistema di riferimento dato”.

Ricordiamo adesso un po' di terminologia legata alle coordinate cartesiane. Supponiamo di aver fissato un sistema di riferimento $R(O, A_1, A_2)$. L'insieme dei punti con ordinata positiva (rispettivamente, negativa) sarà detto *semipiano superiore*

(rispettivamente, *semipiano inferiore*); l'insieme dei punti con ascissa positiva (rispettivamente, negativa) sarà detto *semipiano destro* (rispettivamente, *semipiano sinistro*). L'insieme dei punti con ascissa e ordinata positiva sarà detto *primo quadrante*; quello dei punti con ascissa negativa e ordinata positiva *secondo quadrante*; quello dei punti con ascissa e ordinata negative *terzo quadrante*; quello dei punti con ascissa positiva e ordinata negativa *quarto quadrante*. In altre parole, i quadranti sono ordinati in senso antiorario, partendo da quello con ascisse e ordinate positive, che è la porzione di piano che si usa più spesso.

3.3 Equazioni e disequazioni

Oltre a permettere di rappresentare visivamente i grafici di funzioni, le coordinate cartesiane servono anche a dare una rappresentazione grafica di equazioni e disequazioni. *In questa sezione supporremo di aver fissato una volta per tutte un sistema di riferimento.*

Cominciamo con un'osservazione molto semplice. La costruzione che abbiamo fatto mostra (perché?) che tutti i punti di una retta parallela all'asse delle ascisse hanno la stessa ordinata, e che tutti i punti con la stessa ordinata stanno su una retta parallela all'asse delle ascisse⁵. In altre parole, per ogni $c \in \mathbb{R}$ l'insieme dei punti (vedi l'Osservazione 3.21)

$$\{(x, c) \mid x \in \mathbb{R}\} = \{(x, y) \in \mathbb{R}^2 \mid y = c\} \subset \mathbb{R}^2 \quad (3.3)$$

è una retta parallela all'asse delle ascisse; e, viceversa, ogni retta parallela all'asse delle ascisse è descrivibile in questo modo. Molto spesso, si dice che la retta $\{(x, y) \in \mathbb{R}^2 \mid y = c\}$ è *data dall'equazione* $y = c$, o, ancora, che è *la retta di equazione* $y = c$, intendendo che è l'insieme dei punti (x, y) del piano che soddisfano la relazione $y = c$. Per esempio, l'asse delle ascisse è la retta di equazione $y = 0$.

Ragionando nello stesso modo vediamo che ogni retta parallela all'asse delle ordinate è descritta da un'equazione della forma $x = c$ con $c \in \mathbb{R}$, e viceversa che gli insiemi del piano di equazione $x = c$ sono tutti rette parallele all'asse delle ordinate. In particolare, l'asse delle ordinate è esattamente la retta di equazione $x = 0$.

Possiamo descrivere in modo analogo il grafico Γ_f di una funzione $f: I \rightarrow \mathbb{R}$ reale di variabile reale con dominio $I \subseteq \mathbb{R}$. Infatti, sappiamo che

$$\Gamma_f = \{(x, y) \in I \times \mathbb{R} \mid y = f(x)\} ;$$

quindi possiamo dire che il grafico di f è l'insieme di equazione $y = f(x)$. Nota che stiamo sottintendendo il fatto che ci limitiamo a considerare punti (x, y) con x appartenente al dominio di f , perché altrimenti la scrittura $f(x)$ non avrebbe senso.

⁵ Convinciti che queste ultime due frasi dicono cose diverse, e chiediti perché abbiamo bisogno di entrambe per l'affermazione seguente.

ESEMPIO 3.34 Sia $f: \mathbb{R} \rightarrow \mathbb{R}$ il polinomio quadratico $f(x) = x^2 - 3x + 2$. Allora il grafico di f è l'insieme di equazione $y = x^2 - 3x + 2$, che è una parabola. Siccome la funzione f è definita su tutta la retta reale, non ci sono limitazioni sui valori che x può assumere.

ESEMPIO 3.35 Sia $f: [-1, 1] \rightarrow \mathbb{R}$ la funzione $f(x) = \sqrt{1 - x^2}$. Allora il grafico di f è l'insieme di equazione $y = \sqrt{1 - x^2}$, cioè la semicirconferenza superiore di centro l'origine e raggio 1 (e ovviamente x è compreso nell'intervallo $[-1, 1]$, perché la funzione f è definita solo in quell'intervallo).

Supponiamo ora di voler risolvere l'equazione $f(x) = 0$, cioè di voler trovare gli elementi del dominio della funzione $f: I \rightarrow \mathbb{R}$ che vengono mandati in 0 da f . Per capire il significato geometrico del problema, scomponiamo l'equazione $f(x) = 0$ nelle due equazioni $f(x) = y$ e $y = 0$; chiaramente, il numero x soddisfa l'equazione $f(x) = 0$ se e solo se la coppia (x, y) soddisfa le due equazioni $f(x) = y$ e $y = 0$. Ma $f(x) = y$ è l'equazione del grafico di f , e $y = 0$ è l'equazione dell'asse delle ascisse; quindi i punti (x, y) che soddisfano entrambe queste equazioni sono quelli che stanno sull'intersezione del grafico di f con l'asse delle ascisse. Quindi *le soluzioni dell'equazione $f(x) = 0$ sono le ascisse dei punti di intersezione del grafico di f con l'asse delle ascisse*. In maniera assolutamente analoga si vede che, più in generale, le soluzioni dell'equazione $f(x) = c$ sono le ascisse dei punti di intersezione del grafico di f con la retta di equazione $y = c$.

ESEMPIO 3.36 Se $f(x) = x^2 - 3x + 2$ si vede subito che il grafico di f interseca l'asse delle ascisse nei punti $(1, 0)$ e $(2, 0)$. Siccome il grafico di f è una parabola rivolta verso l'alto, ha ordinata minima nel vertice $(1/2, f(1/2)) = (1/2, -1/4)$. Quindi se $c < -1/4$ il grafico di f non interseca la retta $y = c$, e di conseguenza l'equazione $f(x) = c$ non ha soluzione. La retta $y = -1/4$ interseca il grafico in un solo punto, per cui l'equazione $f(x) = -1/4$ ha un'unica soluzione. Infine, se $c > -1/4$ la retta $y = c$ interseca il grafico di f in due punti, per cui l'equazione $f(x) = c$ ha due soluzioni distinte.

La retta $y = c$ è il grafico della funzione costante $g(x) \equiv c$. Se consideriamo una funzione $g: J \rightarrow \mathbb{R}$ qualsiasi, con $J \subseteq \mathbb{R}$, allora un ragionamento analogo al precedente mostra che *le soluzioni dell'equazione $f(x) = g(x)$ sono le ascissi dei punti di intersezione del grafico di f con il grafico di g* .

Osservazione 3.22 Questo tipo di interpretazione geometrica *non* serve (quasi mai) a trovare esplicitamente la soluzione di un'equazione. Serve invece a dare un'idea qualitativa della struttura dell'insieme delle soluzioni (quante dovrebbero essere, in che zona si dovrebbero trovare, eccetera), e a visualizzare cosa sta succedendo.

Portiamo questo procedimento ancora oltre. C'è ancora un altro modo per descrivere il grafico di una funzione: come l'insieme dei punti (x, y) che risolvono l'equazione $f(x) - y = 0$ (e con x nel dominio di f). In altre parole, stiamo prendendo la funzione di *due* variabili reali $F(x, y) = f(x) - y$ e considerando l'insieme dei punti (x, y) che soddisfano l'equazione $F(x, y) = 0$.

Chiaramente, possiamo ripetere l'operazione con qualsiasi funzione reale di due variabili reali. In altre parole, data una funzione $F: A \rightarrow \mathbb{R}$, con $A \subseteq \mathbb{R}^2$, possiamo considerare l'insieme *di equazione* $F(x, y) = 0$, cioè l'insieme

$$S = \{(x, y) \in A \mid F(x, y) = 0\} \subseteq \mathbb{R}^2 ;$$

spesso scriveremo anche $S = \{F(x, y) = 0\}$.

Usare le equazioni è una delle tecniche più efficienti per descrivere sottoinsiemi del piano, anche molto più generali dei grafici.

ESEMPIO 3.37 La circonferenza di centro l'origine e raggio unitario chiaramente non è un grafico, in quanto ci sono rette verticali che la intersecano in due punti; ricordando l'Esempio 3.35, possiamo solo dire (perché?) che è l'unione di due grafici. Usando la tecnica appena introdotta possiamo descriverla con un'unica equazione: è l'insieme di equazione $x^2 + y^2 - 1 = 0$, o, anche, di equazione $x^2 + y^2 = 1$.

CURIOSITÀ 3.2 Tramite questa tecnica l'introduzione delle coordinate cartesiane permette di studiare insiemi geometrici molto più generali di quelli alla portata dei geometri greci, limitati a rette, circonferenze e poco altro. Un esempio di risultato completamente fuori dalla portata dei geometri classici è il *teorema di Bezout*. Supponiamo di avere due insiemi del piano S_1 ed S_2 , descritti rispettivamente dalle equazioni $F_1(x, y) = 0$ e $F_2(x, y) = 0$, dove F_1 ed F_2 sono due polinomi in due variabili di grado rispettivamente d_1 e d_2 . Allora Bezout ci dice che i due insiemi S_1 ed S_2 si intersecano in al più $d_1 d_2$ punti, o in un numero infinito di punti (nel qual caso i due polinomi hanno un fattore comune). Per esempio, questo ci dice che se una retta S_2 (che è data da un'equazione di grado 1) interseca l'insieme S_1 di equazione $F_1(x, y) = 0$ in più di d_1 punti allora è tutta contenuta in S_1 .

Osservazione 3.23 Non è difficile trovare l'equazione dell'unione e dell'intersezione di insiemi definiti tramite un'equazione. Prendiamo $S_1 = \{F_1(x, y) = 0\}$ e $S_2 = \{F_2(x, y) = 0\}$. Siccome il prodotto di due numeri è nullo se e solo se almeno uno dei due lo è, otteniamo (perché?)

$$S_1 \cup S_2 = \{F_1(x, y) F_2(x, y) = 0\} ,$$

cioè l'unione $S_1 \cup S_2$ ha equazione $F_1 F_2 = 0$. Analogamente, siccome la somma dei quadrati di due numeri reali è zero se e solo se entrambi sono nulli, otteniamo (perché?)

$$S_1 \cap S_2 = \{F_1(x, y)^2 + F_2(x, y)^2 = 0\} ,$$

cioè l'intersezione $S_1 \cap S_2$ ha equazione $F_1^2 + F_2^2 = 0$.

Anche le disequazioni hanno un'analogia interpretazione grafica. Per esempio, il semipiano superiore è rappresentato dalla disequazione $y > 0$: infatti, un punto (x, y) appartiene al semipiano superiore se e solo se ha ordinata positiva. Ragionando come abbiamo fatto finora (controlla) si vede quindi che *le soluzioni della disequazione $f(x) > 0$ sono le ascisse dei punti del grafico di f contenuti nel semipiano superiore*.

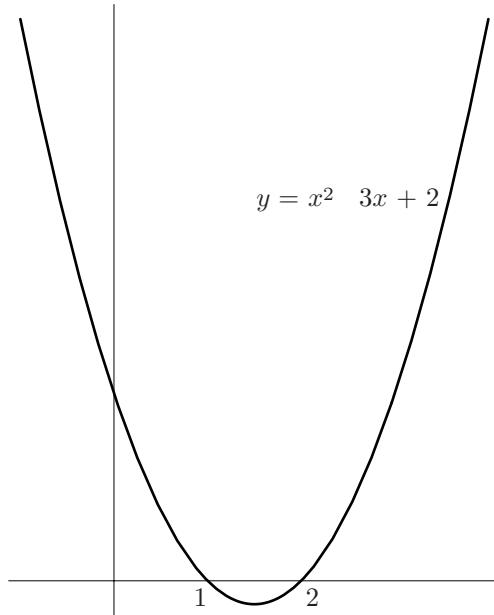


Figura 3.8

ESEMPIO 3.38 Vogliamo risolvere la disequazione $x^2 - 3x + 2 > 0$. Poniamo $f(x) = x^2 - 3x + 2$; risolvere la disequazione equivale (perché?) a determinare $f^{-1}(\mathbb{R}^+)$. Il grafico di f è una parabola (vedi la Figura 3.8), con la concavità rivolta verso l'alto in quanto il coefficiente direttivo è positivo (è uguale a 1). Abbiamo già osservato (Esempio 3.36) che il grafico di f interseca l'asse delle ascisse nei punti $(1, 0)$ e $(2, 0)$. Siccome la concavità è diretta verso l'alto, il grafico di f rimane nel semipiano positivo a destra di $(2, 0)$ e a sinistra di $(1, 0)$; quindi l'insieme delle soluzioni della disequazione è dato dall'unione delle due semirette $(-\infty, 1)$ e $(2, +\infty)$:

$$f^{-1}(\mathbb{R}^+) = \{x \in \mathbb{R} \mid x < 1 \text{ o } x > 2\} = (-\infty, 1) \cup (2, +\infty).$$

Concludiamo questa sezione con un esempio di una situazione che si verifica spesso nella pratica scientifica, coinvolgente diverse equazioni e disequazioni lineari, e la cui rappresentazione grafica aiuta sensibilmente nel suggerire la soluzione.

ESEMPIO 3.39 *Al tuo laboratorio è stato affidato il compito di contare il numero di batteri presenti in 1000 campioni, e devi concludere il lavoro in al massimo 80 ore. Hai a disposizione due possibili contatori. Il primo è più semplice, lo puoi affidare anche al tuo assistente, e in un'ora conta il numero di batteri presenti in 4 campioni. Il secondo contatore è invece un modello più sofisticato e in un'ora conta il numero di batteri presenti in 10 campioni; ma per farlo funzionare hai bisogno di un tecnico specializzato, che ti costa 15 euro all'ora. Siccome si avvicina il suo compleanno, hai deciso di pagare anche il tuo assistente per questo lavoro;*

ma solo 8 euro all'ora, perché il suo contatore è più semplice da usare. Come dividi i campioni fra i due contatori in modo da spendere il meno possibile? Cominciamo col cercare di capire quali sono i vincoli imposti dal problema. Indichiamo con x_1 il numero di ore di funzionamento del primo contatore, e con x_2 il numero di ore di funzionamento del secondo contatore; il tuo obiettivo è trovare la coppia (x_1, x_2) che ti permetta di contare tutti i campioni spendendo il meno possibile. Siccome devi concludere il conteggio entro 80 ore (e chiaramente le ore sono numeri positivi) otteniamo i vincoli

$$0 \leq x_1 \leq 80 \quad \text{e} \quad 0 \leq x_2 \leq 80 .$$

Dunque la soluzione (x_1, x_2) deve trovarsi nella regione di piano definita da queste disequazioni, che è chiaramente un quadrato Q contenuto nel primo quadrante.

Ma non basta. Tu devi contare 1000 campioni, il primo contatore ne tratta 4 all'ora, il secondo 10 all'ora; quindi devi scegliere x_1 ed x_2 in modo da avere

$$4x_1 + 10x_2 = 1000 ,$$

o, equivalentemente,

$$x_2 = 100 - \frac{2}{5}x_1 .$$

Dunque la soluzione (x_1, x_2) deve appartenere anche al grafico del polinomio lineare $p(x_1) = 100 - \frac{2}{5}x_1$, che è una retta; quindi deve trovarsi nel segmento ottenuto intersecando questa retta con il quadrato Q . L'estremo superiore P_1 di questo segmento è l'intersezione del grafico di p con la retta $x_2 = 80$; quindi (perché?) $P_1 = (50, 80)$. Analogamente, l'estremo destro P_2 di questo segmento è l'intersezione del grafico di p con la retta $x_1 = 80$; quindi (perché?) $P_2 = (80, 68)$. In altre parole, la soluzione deve stare sul pezzo del grafico di p relativo a x_1 che varia da 50 a 80; vedi la Figura 3.9.

Calcoliamo ora il costo. Far lavorare il primo contatore per x_1 ore e il secondo contatore per x_2 ore ti costa

$$C(x_1, x_2) = 8x_1 + 15x_2$$

euro. Quindi il tuo obiettivo è trovare il punto (x_1, x_2) appartenente al segmento di estremi P_1 e P_2 per cui il costo $C(x_1, x_2)$ sia il minimo possibile.

Per risolvere questo problema, notiamo prima di tutto che la soluzione deve appartenere al grafico del polinomio p , per cui dev'essere della forma $(x_1, p(x_1))$. Sostituendo allora $p(x_1)$ a x_2 nella formula del costo troviamo

$$C(x_1, p(x_1)) = 8x_1 + 15 \left(100 - \frac{2}{5}x_1 \right) = 1500 + 2x_1 .$$

Questa è una funzione lineare, con coefficiente di x_1 positivo; quindi cresce al crescere di x_1 . Il costo minimo si ottiene quindi per il valore minimo di x_1 , che abbiamo visto essere $x_1 = 50$ e corrisponde a $x_2 = 80$. Il costo minimo è dunque $C(50, 80) = 1600$ euro. Riassumendo, *il costo minimo di 1600 euro si ottiene*

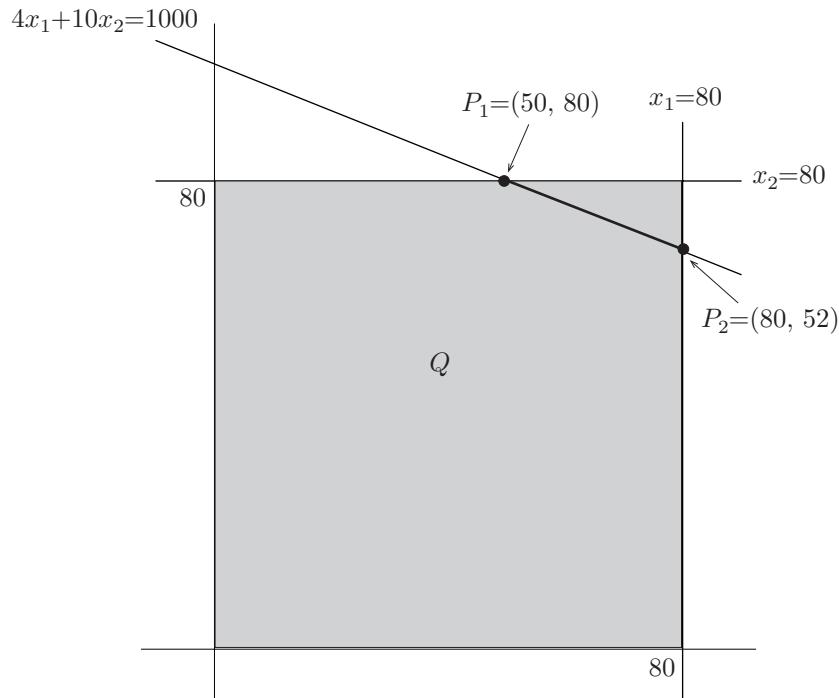


Figura 3.9 Programmazione lineare.

facendo contare 200 campioni in 50 ore al primo contatore, e 800 campioni in 80 ore al secondo contatore. In particolare, contrariamente a quanto forse ci si sarebbe potuti aspettare, la soluzione più economica si ottiene facendo lavorare di più il contatore più costoso, e non quello meno costoso: il motivo è che il secondo contatore è sensibilmente più efficiente del primo (conta in un'ora più del doppio dei campioni costando meno del doppio). Infine, nota che la soluzione si trova sul bordo dell'insieme determinato dai vincoli.

L'esempio precedente è un tipico problema di *programmazione lineare*: si vuole trovare il minimo (o il massimo) di una funzione lineare ristretta a un insieme definito da un numero finito di equazioni o disequazioni lineari. Si può dimostrare che la soluzione è sempre un punto sul bordo dell'insieme; e sono stati sviluppati algoritmi anche molto sofisticati per determinarla in maniera efficiente.

3.4 Diagrammi cartesiani

In questa e nella prossima sezione vogliamo riassumere le modalità più comuni con cui rappresentare graficamente dei dati.

Il primo modo è tramite un diagramma cartesiano.

ESEMPIO 3.40 Hai convinto il tuo assistente a misurarsi il livello di glucosio nel sangue una volta all'ora per un'intera giornata; i risultati sono riportati nella

Ora	Livello di glucosio (mmol/l)	Ora	Livello di glucosio (mmol/l)
0	5.52	12	5.11
1	6.48	13	6.90
2	6.06	14	6.55
3	5.12	15	5.91
4	4.32	16	5.08
5	4.81	17	5.23
6	5.75	18	6.11
7	6.66	19	6.50
8	6.12	20	6.06
9	5.09	21	5.33
10	4.54	22	4.43
11	4.67	23	5.05

TABELLA 3.1

Tabella 3.1.

Abbiamo quindi 24 coppie di dati: da (0, 5.52) a (23, 5.05). Per rappresentarli graficamente, fissiamo un sistema di riferimento cartesiano sul piano, riportando sull'asse delle ascisse i tempi e sull'asse delle ordinate il livello di glucosio. Ovviamen-
te conviene usare un sistema dimetlico, in quanto gli intervalli di variazione (e le unità di misura) dei tempi e dei livelli di glucosio sono alquanto diversi. In questo modo a ogni coppia di dati corrisponde un punto del piano, e otteniamo la Figura 3.10, che rende molto meglio l'andamento del livello di glucosio del tuo assistente di quanto possa farlo la mera lista di dati della Tabella 3.1 (e rivela che il tuo assistente si è fatto uno spuntino di mezzanotte).

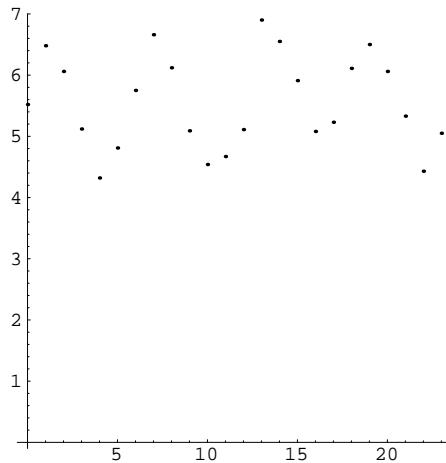


Figura 3.10 Diagramma cartesiano.

In generale, la rappresentazione grafica tramite un diagramma cartesiano è utile quando si devono visualizzare *dati numerici dipendenti da una variabile numerica*.

Sulle ascisse si mettono i possibili valori della variabile indipendente (tempo, quantità di sostanza, energia d'irraggiamento, ecc.) di solito controllata direttamente dallo sperimentatore, mentre le ordinate rappresentano i valori dei dati forniti dall'esperimento.

Osservazione 3.24 Un diagramma cartesiano può essere usato anche per altri tipi di coppie di dati. Per esempio, possiamo aver misurato apertura alare e peso di una popolazione di colibrì composta da 24 individui. Otteniamo 24 coppie di dati, che possiamo visualizzare in un diagramma cartesiano. In questo caso, la scelta di quale variabile usare come ascissa (se il peso o l'apertura alare) è del tutto lasciata allo sperimentatore.

Anche se non è necessariamente sempre vero, la rappresentazione tramite un diagramma cartesiano si usa soprattutto quando i dati sono delle *variabili continue*, cioè variabili che possono assumere (a priori) tutti i valori intermedi fra quelli effettivamente misurati.

ESEMPIO 3.41 Per stressare ulteriormente il tuo assistente, avresti potuto chiedergli di misurarsi il livello di glucosio nel sangue ogni mezz'ora, od ogni dieci minuti, od ogni dieci secondi; e avrebbe trovato valori intermedi rispetto a quelli rappresentati nella Figura 3.10.

Uno dei problemi principali che uno scienziato si trova ad affrontare è proprio quello dell'*interpolazione*: trovare un modo affidabile per predirre i valori intermedi della quantità studiata a partire dai dati effettivamente misurati. Un problema collegato (ma molto più delicato) è quello dell'*estrapolazione*: predirre il valore della quantità studiata per valori della variabile indipendente al di fuori dell'intervallo misurato nell'esperimento.

ESEMPIO 3.42 Nel caso del livello di glucosio del tuo assistente, interpolare significa predirre il livello di glucosio in altri istanti della stessa giornata (per esempio, alle 12:30, oppure alle 02:45); estrapolare significa invece predirre il suo livello di glucosio in giorni diversi da quello in cui è stato condotto l'esperimento (per esempio, il giorno dopo o un mese prima). È chiaro che l'estrapolazione è molto più rischiosa dell'interpolazione, e le predizioni ottenute sono da usare con precauzione.

Il tipo di interpolazione più semplice consiste nell'unire con dei segmenti i punti del diagramma, ottenendo un *diagramma a spezzate* come quello della Figura 3.11.

Questo metodo serve a visualizzare meglio l'andamento dei dati (se crescono o decrescono, per esempio), ma per il resto ha molti difetti. Per esempio, ciascun segmento dipende solo dai suoi estremi e segue una legge propria, senza tenere conto del resto dei dati e senza tentare di trovare una singola legge che descriva tutti i dati disponibili. Inoltre, è rigorosamente inutile per l'estrapolazione, e non è applicabile se ci sono più dati con la stessa ascissa.

Sono chiaramente necessarie tecniche di interpolazione più sofisticate. Nel prossimo capitolo vedremo come interpolare una retta fra i dati (*metodo dei minimi*

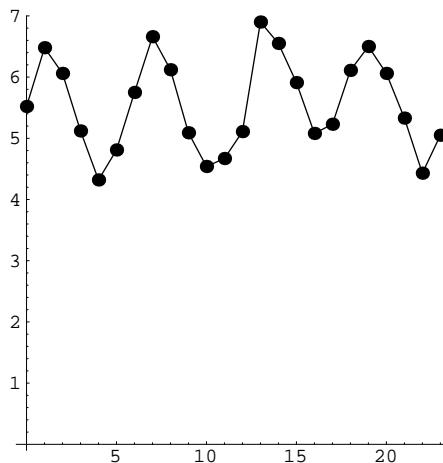


Figura 3.11 Diagramma a spezzate.

quadrati), come misurare la bontà dell’interpolazione fornita dalla retta, e accenneremo a tecniche di interpolazione più complesse (utili per dati dall’andamento periodico quale sembrerebbe essere il livello di glucosio nel sangue del tuo assistente).

Osservazione 3.25 Nella maggior parte delle situazioni sperimentali i dati sono misurati con un certo margine di errore, e indicati con espressioni del tipo $v \pm e$, come abbiamo visto nel Capitolo 1. Un modo per rappresentare graficamente questa incertezza nei diagrammi cartesiani in cui la variabile indipendente è nota con precisione (per esempio perché scelta dallo sperimentatore, come nel caso dell’ora della misura del livello di glucosio del tuo assistente) mentre la variabile dipendente è misurata con un certo errore, consiste nel sostituire i punti con dei segmenti verticali centrati nel valore stimato e di altezza il doppio dell’errore assoluto. Un esempio nel caso del livello di glucosio del tuo assistente è mostrato nella Figura 3.12.

3.5 Istogrammi

Un’altra situazione che si presenta comunemente è quella di dati dipendenti da variabili qualitative. Per rappresentare questi dati si usa un *istogramma*: un diagramma contenente per ciascun valore della variabile qualitativa una colonna di altezza pari al dato corrispondente a quel valore della variabile.

ESEMPIO 3.43 Possiamo riportare su un istogramma i pesi delle cavie dell’Esempio 2.19. La variabile indipendente è l’etichetta delle cavie (che in questo esempio è un numero da 1 a 15, ma poteva benissimo essere una lettera o il nome delle cavie), e la variabile dipendente, che determina l’altezza delle colonne, è il peso della cavia corrispondente. Otteniamo la Figura 3.13.

Un altro caso in cui si usano gli istogrammi è quando i dati dipendono da una

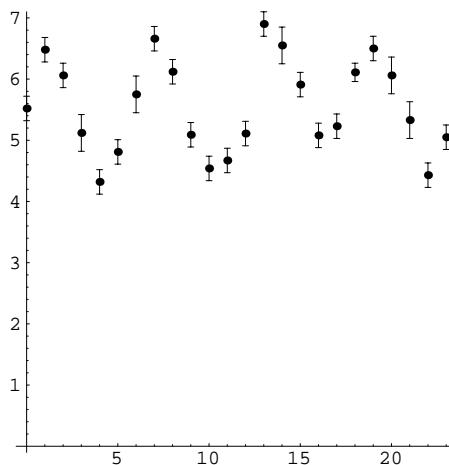


Figura 3.12 Diagramma cartesiano con barre di errore.

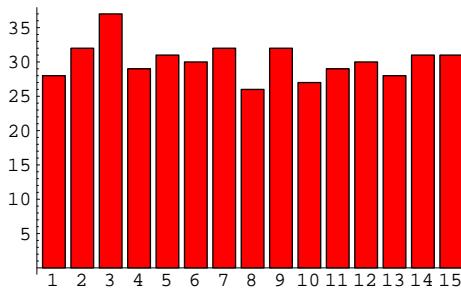


Figura 3.13 Istogramma.

variabile continua *discretizzata*, cioè da una variabile continua considerata però solo in alcuni dei valori possibili.

ESEMPIO 3.44 Un istogramma che rappresenta i valori del livello di glucosio del tuo assistente nel corso di una giornata è riportato nella Figura 3.14.

Non è un caso che le colonne nella Figura 3.14 siano accostate mentre quelle della Figura 3.13 siano separate. L'idea è che nel caso di variabile discretizzata l'istogramma (o meglio, la cima dell'istogramma) rappresenta un'approssimazione del grafico della funzione della variabile continua (un'interpolazione ancora più schematica di quella data dal diagramma a spezzate). Ciascuna colonna rappresenta il comportamento della funzione in tutto un intervallo di valori possibili per la variabile indipendente, l'intervallo coperto dalla base della colonna. A seconda dei casi, l'intervallo è centrato nel valore in cui si è effettuata la misura (come nella Figura 3.14), oppure è situato fra due valori consecutivi in cui si è effettuata la misura (e in tal caso l'altezza corrisponde alla misura effettuata nel valore di sinistra; è il caso della Figura 3.15, sempre riferita al livello del glucosio nel sangue del tuo assistente ma misurato ogni due ore).

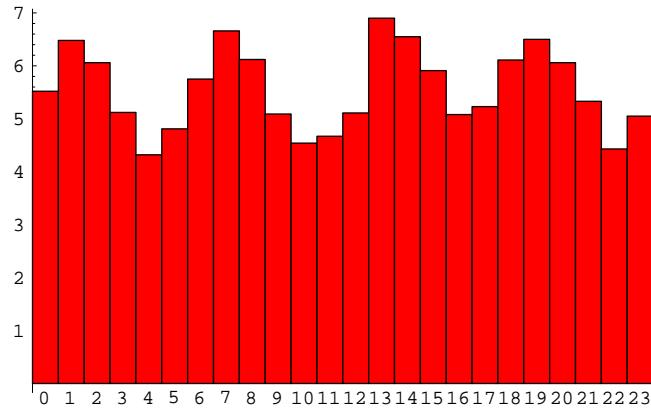


Figura 3.14 Livello di glucosio ogni ora.

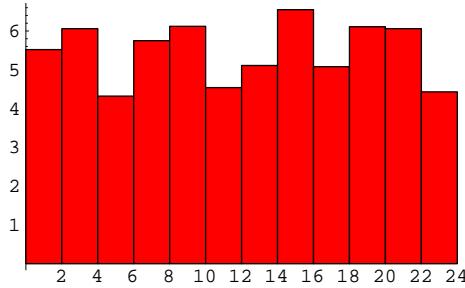


Figura 3.15 Livello di glucosio ogni due ore.

Un altro uso degli istogrammi è per rappresentare i risultati di misure di una stessa quantità (continua) effettuate sugli individui di una popolazione. In tal caso, la variabile “indipendente” rappresentata sull’asse delle ascisse è data dai possibili valori della misura; le colonne dell’istogramma rappresentano invece la frequenza assoluta dei risultati, cioè il numero di campioni che hanno fornito quel particolare risultato. Nota che in questo caso la variabile “dipendente”, la frequenza, è una variabile *discreta*, cioè può assumere solo alcuni valori (i numeri naturali) e non tutti i valori intermedi possibili.

Per costruire l’istogramma, prima di tutto si suddividono i possibili valori della misura in intervalli (anche per poter tenere conto degli eventuali errori di misurazione). Sopra ogni intervallo di valori si costruisce una colonna di *area* proporzionale alla frequenza assoluta, cioè al numero di misure con risultato in quell’intervallo. Se gli intervalli sono tutti della stessa lunghezza, allora l’altezza della colonna rappresenta bene la frequenza (in quanto l’area è il prodotto base per altezza, e la base è la stessa per tutte le colonne); se invece gli intervalli sono di lunghezza diversa è importante che la frequenza sia rappresentata dall’area e non dall’altezza, in quanto altrimenti (come vedrai nell’esempio seguente) si potrebbero ottenere effetti visivamente ingannevoli. In particolare, negli istogrammi di questo genere non va disegnato l’asse delle ordinate.

ESEMPIO 3.45 Vogliamo rappresentare la distribuzione dei pesi delle cavie dell’Esempio 2.19. Abbiamo trovato pesi che variano da 26 a 37 g; una prima rappresentazione naturale è un istogramma con una colonna per ogni peso ottenuto, la cui altezza indichi la frequenza di quel peso. In altre parole, stiamo suddividendo i pesi possibili in intervalli di lunghezza 1 (da 26 g inclusi a 27 g esclusi, cioè $I_1 = [26, 27)$; da 27 g inclusi a 28 g esclusi, cioè $I_2 = [27, 28)$; eccetera) e mettendo su ciascun intervallo una colonna di altezza (o area, visto che gli intervalli hanno tutti la stessa lunghezza) proporzionale al numero di cavie con peso contenuto in quell’intervallo. Otteniamo la Figura 3.16.

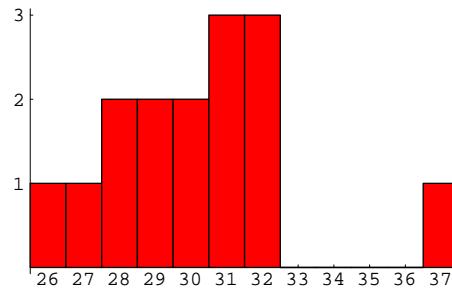


Figura 3.16 Frequenza dei pesi delle cavie.

Se invece prendiamo gli intervalli $(-\infty, 26)$, $[26, 29)$, $[29, 31)$, $[31, 33)$, $[33, 38)$, e $[38, +\infty)$ e rappresentiamo la frequenza con l’area otteniamo la Figura 3.17. Se avessimo rappresentato la frequenza con l’altezza avremmo ottenuto l’istogramma della Figura 3.18, in cui sembrano esserci molte più cavie con peso fra 33 e 38 g di quante ce ne siano in realtà.

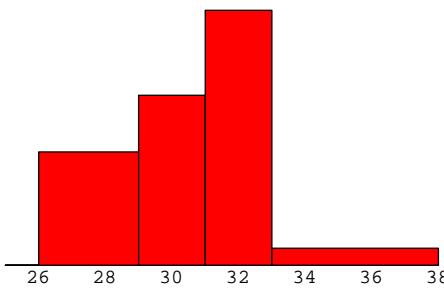


Figura 3.17 Frequenza dei pesi delle cavie con intervalli di lunghezza variabile.

Concludiamo questa sezione ricordando un altro tipo di rappresentazione grafica, gli *aerogrammi* (o *diagrammi a torta*), utile quando si vogliono mostrare dati percentuali, con l’obiettivo di visualizzare principalmente le dimensioni relative.

Supponiamo di avere dei dati percentuali p_1, \dots, p_n con $p_1 + \dots + p_n = 100\%$. L’areogramma che rappresenta questi dati consiste in un cerchio suddiviso in n settori circolari di ampiezza rispettivamente $(p_1/100) \cdot 360^\circ, \dots, (p_n/100) \cdot 360^\circ$ in modo da riempire l’intero disco. Per esempio, la Figura 3.19 contiene l’aerogramma dei dati riportati nella prima metà della Tabella 2.7.

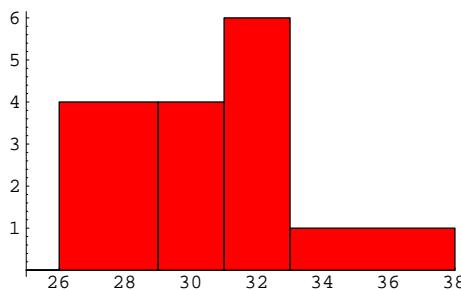


Figura 3.18 Frequenza data dall'altezza.

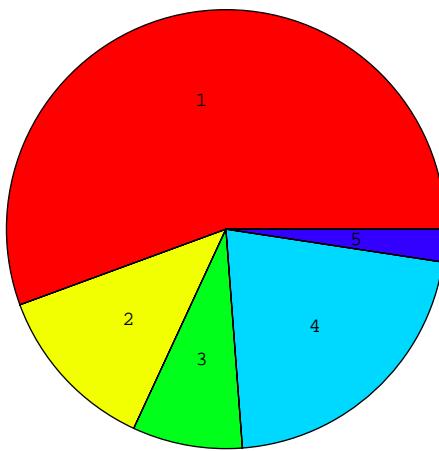


Figura 3.19 Aerogramma.

3.6 Media, mediana e moda

In molte situazioni sperimentali è necessario riassumere molti dati con un solo numero. Vediamo due esempi tipici.

ESEMPIO 3.46 Supponiamo di voler misurare la lunghezza dell'assone della cellula neurale di un ratto. Ripetiamo la misura dieci volte, ottenendo i seguenti risultati, espressi in micrometri:

70.78, 74.22, 74.03, 71.71, 70.97, 73.47, 69.28, 69.62, 72.31, 72.76 .

La differenza fra i vari risultati è dovuta agli (inevitabili) errori di misurazione; il nostro obiettivo è estrarre da queste misure un unico valore per la lunghezza dell'assone, scelto in modo che sia plausibile che i risultati delle nostre misure siano variazioni casuali attorno a questo valore.

ESEMPIO 3.47 Nell'Esempio 2.19 abbiamo misurato il peso di 15 cavie, ottenendo

i seguenti pesi in grammi:

$$28, 32, 37, 29, 31, 30, 32, 26, 32, 27, 29, 30, 28, 31, 31.$$

Vorremmo trovare un numero che rappresenti il peso tipico di questa popolazione di cavie.

In entrambi questi casi abbiamo un numero finito di dati x_1, \dots, x_n dello stesso tipo (molte misure della stessa quantità, sia sullo stesso campione sia su campioni diversi), e vorremmo trovare un numero \bar{x} che riassuma questi dati.

Supponiamo di riassumere i nostri dati con il valore \bar{x} . Se $x_i = \bar{x}$, il dato i -esimo è perfettamente rappresentato da \bar{x} . Ma se $x_i \neq \bar{x}$, il valore \bar{x} rappresenta il dato x_i solo a meno dell'*errore* (o *scarto*) $\bar{x} - x_i$. Quindi è ragionevole pensare che la scelta migliore di \bar{x} sia quella che *minimizza* gli errori $\bar{x} - x_i$.

Ma gli errori sono a priori tanti, uno per ogni dato, e non possiamo pretendere di minimizzarli tutti indipendentemente l'uno dell'altro; dobbiamo cercare di minimizzarli nel loro complesso.

Una prima possibilità consiste nel cercare di minimizzare la somma

$$\sum_{i=1}^n (\bar{x} - x_i)$$

degli errori, o, più precisamente, il suo valore assoluto. Siccome il numero con valore assoluto minimo è zero, vogliamo vedere se esiste un numero $\bar{x} \in \mathbb{R}$ tale che $\sum_{i=1}^n (\bar{x} - x_i) = 0$. Ma

$$\sum_{i=1}^n (\bar{x} - x_i) = (\bar{x} - x_1) + \dots + (\bar{x} - x_n) = n\bar{x} - (x_1 + \dots + x_n) = n\bar{x} - \sum_{i=1}^n x_i,$$

e quindi

$$\sum_{i=1}^n (\bar{x} - x_i) = 0 \iff \bar{x} = \frac{x_1 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Il numero \bar{x} così ottenuto si chiama *media* (o *media aritmetica* o *valor medio*) dei dati x_1, \dots, x_n :

$$\bar{x} = \text{Media}(x_i) = \frac{1}{n} \sum_{i=1}^n x_i.$$

ESEMPIO 3.48 La media delle misure delle lunghezze dell'assone del ratto dell'Esempio 3.46 è

$$\begin{aligned} \bar{x} &= \frac{70.78 + 74.22 + 74.03 + 71.71 + 70.97 + 73.47 + 69.28 + 69.62 + 72.31 + 72.76}{10} \\ &= 71.915. \end{aligned}$$

ESEMPIO 3.49 La media dei pesi delle cavie dell'Esempio 3.47 è

$$\begin{aligned}\bar{x} &= \frac{28 + 32 + 37 + 29 + 31 + 30 + 32 + 26 + 32 + 27 + 29 + 30 + 28 + 31 + 31}{15} \\ &= 30.2.\end{aligned}$$

Osservazione 3.26 Gli errori $\bar{x} - x_i$ si suddividono naturalmente in due categorie: quelli positivi (per cui $x_i < \bar{x}$) e quelli negativi (per cui $x_i > \bar{x}$). La media è l'unico $\bar{x} \in \mathbb{R}$ per cui gli errori negativi bilanciano esattamente quelli positivi.

Osservazione 3.27 Potresti obiettare che in un certo senso stiamo barando: gli errori sono errori indipendentemente dal segno, per cui dovremmo piuttosto cercare di minimizzare quantità del tipo

$$\sum_{i=1}^n (\bar{x} - x_i)^2, \quad (3.4)$$

in cui ogni errore non nullo contribuisce con un termine positivo. Ma puoi stare tranquilla: come vedremo nel prossimo capitolo, la media minimizza anche (3.4).

Osservazione 3.28 La media risente della presenza di valori estremi nei dati: un solo dato di valore sensibilmente diverso dagli altri può spostare la media in modo significativo (vedi l'esempio successivo). Per questo motivo la presenza di dati anomali nei risultati di un esperimento va esaminata con attenzione: il valore medio calcolato escludendo questi dati estremi può essere più significativo del valore medio calcolato includendoli, *a meno che* la loro presenza non indichi l'esistenza di un fenomeno imprevisto (e quindi interessante) comparso solo in quelle misure. In ogni caso, più i dati sono distribuiti in modo uguale su entrambi i lati della media (cioè, più la media si avvicina alla mediana; vedi sotto) più la media è rappresentativa dei dati nel loro complesso. Fortunatamente, come discuteremo in un prossimo capitolo, variazioni nei dati prodotte da errori casuali soddisfano questa condizione.

ESEMPIO 3.50 Guardando i pesi delle cavie dell'Esempio 3.47 si nota subito la presenza di un valore anomalo, i 37 g di peso della terza cavia. Se escludiamo questo valore, la media dei pesi diventa

$$\begin{aligned}\bar{x} &= \frac{28 + 32 + 29 + 31 + 30 + 32 + 26 + 32 + 27 + 29 + 30 + 28 + 31 + 31}{14} \\ &\simeq 29.71.\end{aligned}$$

Quindi l'esclusione del dato anomalo ha provocato una variazione della media pari a $30.2 - 29.71 = 0.49$ g.

Osservazione 3.29 Nel calcolo della media aritmetica, ogni dato viene contato tutte le volte che compare. Per esempio, nel caso delle cavie i dati assumono 8 valori

distinti (26, 27, 28, 29, 30, 31, 32, 37), e le ripetizioni sono ammesse. Se raggruppiamo insieme i valori ripetuti, possiamo calcolare la media con la formula

$$\bar{x} = \frac{1 \cdot 26 + 1 \cdot 27 + 2 \cdot 28 + 2 \cdot 29 + 2 \cdot 30 + 3 \cdot 31 + 3 \cdot 32 + 1 \cdot 37}{1 + 1 + 2 + 2 + 2 + 3 + 3 + 1}.$$

Più in generale, se abbiamo un gruppo di dati x_1, \dots, x_n che devono essere pesati in maniera diversa (per esempio, perché compaiono con frequenze diverse) si può usare la *media ponderata* (o *pesata*) definita da

$$\bar{x} = \frac{f_1 x_1 + \dots + f_n x_n}{f_1 + \dots + f_n},$$

dove f_1, \dots, f_n sono numeri positivi detti *pesi* dei dati x_1, \dots, x_n . La media aritmetica usuale corrisponde a prendere $f_1 = \dots = f_n = 1$, cioè a pesare tutti i dati nello stesso modo.

CURIOSITÀ 3.3 I pesi hanno una naturale interpretazione probabilistica. Se indichiamo con $F = \sum_{i=1}^n f_i$ la somma delle frequenze e poniamo $p_i = f_i/F$, allora $p_1 + \dots + p_n = 1$, e quindi possiamo interpretare p_i come la probabilità di ottenere il valore x_i facendo la nostra misura.

Usando i p_i possiamo calcolare la media ponderata con la formula

$$\bar{x} = \sum_{i=1}^n p_i x_i. \quad (3.5)$$

Più in generale, in un esperimento probabilistico di spazio degli eventi $\Omega = \{x_1, \dots, x_n\}$ composto da un numero finito di numeri reali la formula (3.5) (dove p_i ora è la probabilità dell'evento semplice x_i) fornisce il *valore atteso* (o *media*) dell'esperimento. Per esempio, il valore atteso della somma del lancio di due dadi a sei facce non truccati è

$$\frac{1}{36} \cdot 2 + \frac{2}{36} \cdot 3 + \frac{3}{36} \cdot 4 + \frac{4}{36} \cdot 5 + \frac{5}{36} \cdot 6 + \frac{6}{36} \cdot 7 + \frac{5}{36} \cdot 8 + \frac{4}{36} \cdot 9 + \frac{3}{36} \cdot 10 + \frac{2}{36} \cdot 11 + \frac{1}{36} \cdot 12 = 7.$$

CURIOSITÀ 3.4 Nel caso di dati *positivi*, potremmo considerare invece degli errori assoluti $x - x_i$ gli errori relativi x/x_i . In questo caso, l'assenza di errore corrisponde ad avere errore relativo uguale a 1; quindi possiamo chiederci se esiste un $\hat{x} \in \mathbb{R}$ per cui il *prodotto* $\prod_{i=1}^n (\hat{x}/x_i)$ sia esattamente uguale a 1. È facile rispondere: siccome

$$\prod_{i=1}^n \frac{\hat{x}}{x_i} = \frac{\hat{x}^n}{\prod_{i=1}^n x_i},$$

il prodotto degli errori relativi è uguale a 1 se e solo se

$$\hat{x} = \left(\prod_{i=1}^n x_i \right)^{1/n}.$$

Questo \hat{x} è detto *media geometrica* di x_1, \dots, x_n . Si può dimostrare che la media geometrica è sempre minore o uguale alla media aritmetica, cioè

$$(x_1 \cdots x_n)^{1/n} \leq \frac{x_1 + \cdots + x_n}{n},$$

con uguaglianza se e solo se $x_1 = \dots = x_n$.

Quando i dati non sono equamente distribuiti attorno alla media, può essere utile riassumerli con un altro valore che soddisfa invece questa condizione: la mediana. Prendiamo i nostri dati e disponiamoli in ordine crescente, in modo da avere

$$x_1 \leq x_2 \leq \dots \leq x_n .$$

Allora la *mediana* dei dati è un numero M tale che esattamente metà dei dati sia minore o uguale a M , e metà sia maggiore o uguale a M : per l'esattezza,

$$M = \begin{cases} x_{(n+1)/2} & \text{se } n \text{ è dispari,} \\ \frac{1}{2}(x_{n/2} + x_{(n/2)+1}) & \text{se } n \text{ è pari.} \end{cases}$$

In altre parole, M è il termine centrale se n è dispari, ed è la media aritmetica dei due termini centrali se n è pari. In particolare, la mediana non è influenzata da dati estremi anomali (ma dipende dal numero dei dati).

ESEMPIO 3.51 Per calcolare la mediana dei pesi delle cavie dell'Esempio 3.47 cominciamo disponendoli in ordine crescente:

$$26 \leq 27 \leq 28 \leq 28 \leq 29 \leq 29 \leq 30 \leq 30 \leq 31 \leq 31 \leq 31 \leq 32 \leq 32 \leq 32 \leq 37 .$$

Abbiamo $n = 15$ dati; quindi la mediana è l'ottavo ($8 = (15 + 1)/2$) dato $M = 30$. Nota che la mediana non cambia anche se aumentiamo il dato estremo 37, in quanto l'ordine degli altri dati non viene modificato. Se togliamo il dato estremo 37, ci rimangono 14 dati, per cui la mediana diventa la media aritmetica fra il settimo ($7 = 14/2$) e l'ottavo ($8 = (14/2) + 1$) dato, per cui $M = (30 + 30)/2 = 30$. Se invece togliamo il dato estremo 26, la mediana diventa la media aritmetica fra il settimo e l'ottavo dato *nel nuovo ordine*, per cui diventa $M = (30 + 31)/2 = 30.5$.

CURIOSITÀ 3.5 La mediana divide i dati in due insiemi con ugual numero di elementi. I valori, ottenuti in modo analogo, che dividono i dati in quattro insiemi con ugual numero di elementi si chiamano *quartili*; in particolare, la mediana è il secondo quartile. Se ci sono moltissimi dati, a volte si usano i *percentili*, che dividono i dati in cento insiemi con ugual numero di elementi — e la mediana è il cinquantesimo percentile.

CURIOSITÀ 3.6 Anche la mediana minimizza un'opportuna somma di errori. Per l'esattezza, si può dimostrare che la mediana M dei dati x_1, \dots, x_n rende minima la somma

$$\sum_{i=1}^n |M - x_i|$$

dei valori assoluti degli errori.

Media e mediana servono per riassumere dati numerici con un unico valore. A volte si vorrebbe operare in modo analogo anche con dati non numerici (colori degli occhi, specie di animale, eccetera). In tal caso si suddividono i dati in classi, e si

chiama *moda* (o *classe modale*) la classe (o le classi, se ce ne sono più d'una) col maggior numero di elementi.

ESEMPIO 3.52 Supponiamo di avere come insieme di dati l'elenco dei maschi italiani morti nel 2002, e suddividiamoli in classi d'età come indicato nella Tabella 2.8. Allora la moda (cioè la classe con più elementi) è la classe 70–79.

ESEMPIO 3.53 Ovviamente, si può calcolare la moda anche di dati numerici: la moda è il dato (o i dati) con frequenza assoluta maggiore. Per esempio, le classi modali dei pesi delle solite cavie sono 31 e 32, in quanto sono i pesi che si ripetono più frequentemente.

3.7 Varianza

La media (o la mediana) non bastano a riassumere ragionevolmente i dati; ci serve anche una misura di quanto la media sia rappresentativa, cioè di quanto i dati si accumulano vicino alla media, o di quanto invece sono sparsi fra tutti i possibili valori. In altre parole, ci serve una misura della dispersione dei dati.

Una prima misura di dispersione è l'*intervallo di variabilità*: la differenza fra il dato massimo x_{\max} e il dato minimo x_{\min} .

ESEMPIO 3.54 Nel caso dei pesi delle cavie, il dato massimo è $x_{\max} = 37$ e il dato minimo è $x_{\min} = 26$, per cui l'intervallo di variabilità è $37 - 26 = 11$.

L'intervallo di variabilità ci dice quanto sono sparsi i dati, ma non quanto sono dispersi rispetto alla media \bar{x} . Inoltre è una misura molto grossolana; per questo motivo è utilizzato di rado. Vogliamo introdurre ora una misura più fine (e molto più utile) della dispersione dei dati intorno alla media.

Siano x_1, \dots, x_n dati di media \bar{x} . Abbiamo già osservato che $\bar{x} - x_i$ indica l'errore che si compie sostituendo la media al posto di x_i (o l'errore che si è compiuto quando misurando \bar{x} si è ottenuto x_i). Per quel che riguarda la dispersione, il segno dell'errore è irrilevante; quindi conviene usare⁶ lo *scarto quadratico* $(\bar{x} - x_i)^2$. La misura di dispersione più comune e più usata è allora la media degli scarti quadratici, chiamata *scarto quadratico medio* o, più spesso, varianza:

$$\text{Var}(x_i) = \frac{1}{n} \sum_{i=1}^n (\bar{x} - x_i)^2. \quad (3.6)$$

⁶ Ci sono due motivi che spingono a usare il quadrato degli errori e non i valori assoluti. Il primo è che il quadrato, essendo un polinomio, ha proprietà algebriche molto più gradevoli della funzione valore assoluto. Il secondo è che la media minimizza la somma dei quadrati degli errori, mentre la somma dei valori assoluti degli errori è minimizzata dalla mediana; vedi l'Osservazione 3.27 e la Curiosità 3.6.

A volte, la varianza si indica anche col simbolo σ^2 . La radice quadrata della varianza si chiama *deviazione standard*, e si indica col simbolo σ o, talvolta, DS:

$$DS(x_i) = \sigma = \sqrt{\text{Var}(x_i)} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\bar{x} - x_i)^2}.$$

Tanto più piccola è la deviazione standard (o la varianza) tanto più i dati sono concentrati attorno al valor medio. In particolare, il *coefficiente di variazione*

$$CV(x_i) = \frac{\sigma}{\bar{x}}$$

è una misura della dispersione dei dati che non dipende dall'unità di misura usata, e che permette di confrontare la dispersione di dati diversi.

Osservazione 3.30 Attenzione a non confondere la varianza con la deviazione standard! Se i dati sono misurati (per esempio) in metri, la varianza è misurata in metri quadri, ed è la deviazione standard a essere misurata in metri. In particolare, è il valore della deviazione standard (e non quello della varianza) che va confrontato con la media.

Osservazione 3.31 Chiaramente, la varianza e la deviazione standard si annullano se e solo se tutti i dati sono uguali. Inoltre, come già anticipato nell'Osservazione 3.27 (e come dimostreremo nel prossimo capitolo), la media \bar{x} è il numero reale rispetto a cui i dati hanno scarto quadratico medio minimo, in quanto minimizza la quantità $\frac{1}{n} \sum_{i=1}^n (\bar{x} - x_i)^2$; anche per questo motivo la media aritmetica è considerata il modo migliore per riassumere i dati.

Osservazione 3.32 C'è un'altra formula per il calcolo della varianza che è spesso utile. Espandendo il quadrato in (3.6) otteniamo

$$\begin{aligned} \text{Var}(x_i) &= \frac{1}{n} \sum_{i=1}^n (\bar{x}^2 - 2\bar{x}x_i + x_i^2) = \frac{1}{n} n\bar{x}^2 - 2\bar{x} \frac{1}{n} \sum_{i=1}^n x_i + \frac{1}{n} \sum_{i=1}^n x_i^2 \\ &= \bar{x}^2 - 2\bar{x}^2 + \frac{1}{n} \sum_{i=1}^n x_i^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 \\ &= \text{Media}(x_i^2) - \text{Media}(x_i)^2, \end{aligned}$$

cioè *la varianza è uguale alla media dei quadrati meno il quadrato della media*.

ESEMPIO 3.55 Calcoliamo la varianza del peso delle cavie dell'Esempio 3.47. Ricordando che la media è 30.2, troviamo che gli scarti quadratici sono

$$4.84, 3.24, 46.24, 1.44, 0.64, 0.04, 3.24, 17.64, 3.24, 10.24, 1.44, 0.04, 4.84, 0.64, 0.64,$$

per cui la varianza è

$$\text{Var} = \frac{1}{15} \left[4.84 + 3.24 + 46.24 + 1.44 + 0.64 + 0.04 + 3.24 + 17.64 + 3.24 + 10.24 + 1.44 + 0.04 + 4.84 + 0.64 + 0.64 \right] = 6.56 ,$$

e la deviazione standard è

$$\sigma = \sqrt{6.56} \simeq 2.56 .$$

Si otteneva lo stesso risultato calcolando la media dei quadrati e sottraendo la media al quadrato:

$$\begin{aligned} \text{Media}(x_i^2) - \text{Media}(x_i)^2 \\ = \frac{1}{15} \left[784 + 1024 + 1369 + 841 + 961 + 900 + 1024 + 676 + 1024 + 729 + 841 + 900 + 784 + 961 + 961 \right] - 30.2^2 \\ = \frac{13779}{15} - 912.04 = 6.56 . \end{aligned}$$

Il coefficiente di variazione dei dati è

$$\text{CV} = \frac{2.56}{30.2} \simeq 0.085 = 8.5\% ,$$

per cui i dati non sono troppo dispersi intorno alla media. Se togliamo il valore estremo 37, la varianza diventa circa 3.73, la deviazione standard circa 1.93, e il coefficiente di variazione circa 6.5%, con un miglioramento del $(8.5 - 6.5)/8.5 = 24\%$ circa.

ESEMPIO 3.56 Calcoliamo la varianza delle misure dell'assone di neurone di ratto dell'Esempio 3.46. Otteniamo

$$\begin{aligned} \text{Var} = \frac{1}{10} \left[70.78^2 + 74.22^2 + 74.03^2 + 71.71^2 + 70.97^2 + 73.47^2 + 69.28^2 + 69.62^2 + 72.31^2 + 72.76^2 \right] - 71.915^2 \\ \simeq 2.751 , \end{aligned}$$

e la deviazione standard è $\sigma \simeq \sqrt{2.751} \simeq 1.66$. In particolare, il coefficiente di variazione è $\text{CV} \simeq 1.66/71.915 \simeq 2.3\%$, per cui i dati sono ben concentrati attorno alla media.

Osservazione 3.33 Come già detto più volte, una delle situazioni sperimentali tipiche in cui è importante il calcolo della media e della deviazione standard è quando si effettuano numerose misure della stessa quantità; le singole misure danno risultati diversi a causa degli errori sperimentali. Gli errori sperimentali, per loro

natura, sono essenzialmente casuali e (per motivi che discuteremo più avanti) si distribuiscono secondo una distribuzione di probabilità detta *distribuzione normale* o *distribuzione Gaussiana* (che definiremo fra un paio di capitoli). Quando questo accade, si può dimostrare che nel 68% circa dei casi il valore vero sarà contenuto nell'intervallo $[\bar{x} - \sigma, \bar{x} + \sigma]$, e che nel 95% circa dei casi il valore vero sarà contenuto nell'intervallo $[\bar{x} - 2\sigma, \bar{x} + 2\sigma]$. Per questo motivo, quando si sa che i dati seguono una distribuzione normale il risultato finale delle misure viene espresso nella forma $\bar{x} \pm 2\sigma$ con una *confidenza* del 95%, per indicare che il risultato vero è contenuto nell'*intervallo di confidenza* $[\bar{x} - 2\sigma, \bar{x} + 2\sigma]$ nel 95% circa dei casi. Si può anche dimostrare che il risultato vero è contenuto nell'intervallo $[\bar{x} - 3\sigma, \bar{x} + 3\sigma]$ nel 99% circa dei casi.

ESEMPIO 3.57 Essendo misure sperimentali dello stesso assone, possiamo assumere che i dati raccolti nell'Esempio 3.46 seguano una distribuzione normale, e dire che la lunghezza dell'assone di neurone è 71.915 ± 3.32 con una confidenza del 95%.

3.8 Campioni e popolazione

Nella maggior parte delle situazioni statistiche e sperimentali, non si ha a disposizione l'intera popolazione, ma solo un campione di essa. Non possiamo misurare le lunghezze degli assi di tutte le cellule neurali di tutti i ratti del mondo (e, probabilmente, neanche di un ratto solo); dobbiamo contentarci di misurarne soltanto alcune, quelle prese da un *campione* dell'intera popolazione.

Partendo dal nostro campione ricaviamo una media e una varianza. La domanda naturale è: questa media e questa varianza rappresentano la media e la varianza dell'intera popolazione? Ci si aspetta che se il campione è davvero scelto a caso ed è composto da un numero sufficientemente alto di individui la risposta sia positiva; in questa sezione cercheremo di vedere se questa aspettativa è corretta o meno.

Cominciamo precisando il problema e fissando alcune notazioni. Noi abbiamo un insieme P , usualmente molto grande, di individui (per esempio, neuroni di ratto) su cui vogliamo effettuare una misura (per esempio, misurare la lunghezza dell'assone). Indichiamo con μ la media e con σ^2 la varianza dei valori ottenuti effettuando la misura su tutta la popolazione; sono valori che di solito non conosciamo, e che vogliamo stimare effettuando le misure su un campione.

Indichiamo con $X_n = \{x_1, \dots, x_n\}$ l'insieme dei valori ottenuti effettuando la misura su un dato campione composto da n individui, e con \mathcal{X}_n la famiglia di tutti gli insiemi X_n ottenuti al variare del campione di n individui nella popolazione. Se $\text{Media}(X_n)$ e $\text{Var}(X_n)$ indicano rispettivamente la media e la varianza dei dati in X_n , ci chiediamo se (e in che senso) $\text{Media}(X_n)$ e $\text{Var}(X_n)$ forniscono una stima sensata di μ e σ^2 .

Cominciamo con la media. Ovviamente non è detto che $\text{Media}(X_n)$ sia uguale a μ ; in generale, ci sarà un errore. Possiamo interpretare $\text{Media}(X_n)$ come una misura di μ , e chiederci qual è la media e la varianza di queste misure al variare di X_n nell'insieme \mathcal{X}_n di tutti i possibili campioni composti da n individui. Alcuni conti un po' elaborati ci forniscono le seguenti risposte:

- (a) la media di $\text{Media}(X_n)$ al variare di $X_n \in \mathcal{X}_n$ è esattamente uguale a μ ;
 (b) la varianza di $\text{Media}(X_n)$ al variare di $X_n \in \mathcal{X}_n$ è uguale a σ^2/n .

Questo vuol dire che le medie su tutti i possibili campioni composti da n elementi hanno media μ , e che queste medie sono sempre più concentrate (meno disperse) attorno a μ al crescere di n , in quanto a patto di prendere n abbastanza grande possiamo rendere σ^2/n piccolo quanto ci pare⁷.

Inoltre, i probabiliti hanno anche dimostrato il seguente risultato, noto come **Teorema del limite centrale**:

- (c) al crescere di n la distribuzione dei valori $\text{Media}(X_n)$ approssima sempre meglio una distribuzione normale di media μ e varianza σ^2/n .

Non siamo ancora in grado (ne ripareremo fra qualche capitolo) di dire cos'è una distribuzione normale, ma nell'Osservazione 3.33 abbiamo già accennato a una importante proprietà di questa distribuzione, proprietà che nel nostro contesto possiamo parafrasare così: *se n è abbastanza grande la media μ della popolazione apparterrà all'intervallo di confidenza $[\text{Media}(X_n) - 2\sigma/\sqrt{n}, \text{Media}(X_n) + 2\sigma/\sqrt{n}]$ per circa il 95% dei campioni composti da n individui*. In altre parole, abbiamo il 95% di probabilità che la media $\text{Media}(X_n)$ di un campione di n individui *scelti a caso* approssimi la media μ dell'intera popolazione a meno di un errore di $2\sigma/\sqrt{n}$. Quindi basta scegliere n abbastanza grande da avere $2\sigma/\sqrt{n}$ molto piccolo e, nel 95% dei casi, $\text{Media}(X_n)$ è una buona approssimazione di μ . Se poi n è così grande che $3\sigma/\sqrt{n}$ è molto piccolo, allora $\text{Media}(X_n)$ diventa una buona approssimazione di μ nel 99% circa dei casi.

Bello, vero? Quasi. C'è ancora un problema da risolvere: non conoscendo σ , come facciamo a sapere se n è abbastanza grande da rendere $2\sigma/\sqrt{n}$ piccolo? Abbiamo bisogno anche di una stima di σ .

La congettura naturale sarebbe che una stima di σ^2 sia data da $\text{Var}(X_n)$. Ma perché questa stima sia utile occorre che valgano delle proprietà analoghe alle proprietà (a) e (b) viste sopra: occorre che la media di $\text{Var}(X_n)$ al variare di $X_n \in \mathcal{X}_n$ sia esattamente σ^2 , e che sia possibile rendere la varianza degli $\text{Var}(X_n)$ arbitrariamente piccola a patto di scegliere n abbastanza grande.

Invece, si può dimostrare il seguente risultato:

- (d) la media di $\text{Var}(X_n)$ al variare di $X_n \in \mathcal{X}_n$ è uguale a $(n - 1)\sigma^2/n$, e non a σ^2 .

Questo risultato suggerisce di introdurre la varianza *campionaria*

$$s^2(X_n) = \frac{1}{n-1} \sum_{i=1}^n (\bar{x} - x_i)^2 = \frac{n}{n-1} \text{Var}(X_n) ,$$

dove $\bar{x} = \text{Media}(X_n)$ e $X_n = \{x_1, \dots, x_n\}$. Allora la proprietà (d) diventa

- (d') la media della varianza campionaria $s^2(X_n)$ al variare di $X_n \in \mathcal{X}_n$ è uguale a σ^2 .

⁷ La quantità σ/\sqrt{n} si chiama errore standard della media.

Inoltre, è anche possibile dimostrare che

- (e) *la varianza della varianza campionaria $s^2(X_n)$ diventa arbitrariamente piccola a patto di scegliere n abbastanza grande.*

Quindi per n abbastanza grande i possibili valori di $s^2(X_n)$ sono ben concentrati attorno a σ^2 , per cui la *deviazione standard campionaria*

$$s(X_n) = \sqrt{s^2(X_n)} = \sqrt{\frac{n}{n-1} \text{DS}(X_n)} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (\bar{x} - x_i)^2}$$

è una buona stima della deviazione standard σ dell'intera popolazione, e quindi $2s(X_n)/\sqrt{n}$ è una buona stima dell'errore standard della media $2\sigma/\sqrt{n}$.

Osservazione 3.34 Avrai certamente notato che in realtà non ci siamo liberati del problema di stabilire quanto grande dev'essere n per essere abbastanza grande: abbiamo ancora bisogno che n sia abbastanza grande perché la varianza della varianza campionaria sia sufficientemente piccola, e perché la distribuzione delle medie sia abbastanza vicina alla distribuzione normale. Senza ulteriori ipotesi sulla popolazione originaria (per esempio, la popolazione originaria segue una distribuzione normale oppure no?) non possiamo dire molto altro; ma nella maggior parte dei casi non è necessario. Infatti, nella maggior parte dei casi lo sperimentatore non può scegliere n liberamente; il numero dei campioni testabili è limitato dal costo, dal tempo disponibile, dalla quantità di materiale... Questa teoria ci fornisce allora un test della qualità del risultato che si è ottenuto. Calcoliamo la media $\text{Media}(X_n)$ e la deviazione standard campionaria $s(X_n)$ sul campione che abbiamo a disposizione; se $2s(X_n)/\sqrt{n}$ è piccolo rispetto a $\text{Media}(X_n)$, cioè se il *coefficiente di variazione campionario*

$$\text{CVC}(X_n) = \frac{s(X_n)}{\sqrt{n} \text{ Media}(X_n)} = \sqrt{\frac{1}{n-1}} \text{CV}(X_n)$$

è piccolo (sensibilmente minore dell'ordine di precisione che vogliamo raggiungere), allora $\text{Media}(X_n)$ è probabilmente una buona approssimazione di μ ; ma se non è piccolo dobbiamo assolutamente allargare il campione.

ESEMPIO 3.58 Supponiamo che le cavie dell'Esempio 3.47 siano solo un campione di una ben più vasta popolazione di cavie; vogliamo vedere se la media dei pesi di questo campione è rappresentativa della media dei pesi dell'intera popolazione. Abbiamo visto che la media è $\bar{x} = 30.2$ g. La deviazione standard era circa 2.56 g; quindi la deviazione standard campionaria è

$$s \simeq \sqrt{\frac{15}{14}} 2.56 \simeq 2.65 \text{ g ,}$$

e il coefficiente di variazione campionario è

$$\text{CVC} \simeq \frac{2.65}{\sqrt{15} \cdot 30.2} \simeq 0.023 = 2.3\% .$$

Quindi il doppio dell'errore standard della media è circa il 4.6% della media \bar{x} . Questo vuol dire che l'intervallo $[\bar{x} - 2s/\sqrt{n}, \bar{x} + 2s/\sqrt{n}]$ ha una lunghezza che è circa il 9.2% del valore centrale, che è un po' alto. Infatti,

$$[\bar{x} - 2s/\sqrt{n}, \bar{x} + 2s/\sqrt{n}] \simeq [28.83, 31.57] ,$$

e quindi non abbiamo una stima precisa su nessuna cifra significativa.

Osservazione 3.35 Vale la pena di ribadire che tutti questi conti suppongono che il campione sia *scelto a caso*; non ci stiamo ponendo il problema di come trovare campioni rappresentativi in qualche senso dell'intera popolazione, e stiamo supponendo che il campione non sia viziato da qualche errore sistematico nella scelta o nella misura.

Osservazione 3.36 Tutti questi ragionamenti si applicano solo nel caso in cui non sia possibile misurare l'intera popolazione; se, al contrario, abbiamo a disposizione l'intera popolazione, calcoliamo direttamente la media μ e la deviazione standard σ senza porci problemi.

Osservazione 3.37 Attenzione a non confondere la deviazione standard con la deviazione standard della media. Un giorno, il tuo assistente ti ha preso da parte e ti ha detto: "Ho effettuato un sacco di misure di una stessa quantità, e poi ho calcolato l'intervallo di confidenza usando la media e la deviazione standard. Ma poi ho pensato: le misure che ho fatto sono solo un campione fra tutte le possibili misure che potrei fare. Dunque la media che ho ottenuto misura la media di tutte le possibili misure a meno della deviazione standard della media; e quindi l'intervallo di confidenza lo dovrei fare usando la media e la deviazione standard della media, e così viene molto più carino perché la deviazione standard della media è molto più piccola della deviazione standard usuale." Come al solito, il tuo assistente si sbaglia. Il punto è che il valore vero della misura *non* è necessariamente la media di tutte le possibili misure; anche nel caso della distribuzione normale, sappiamo solo che cade nell'intervallo di confidenza nel 95% dei casi. La deviazione standard della media ci dice tutt'al più con che precisione possiamo identificare il centro dell'intervallo di confidenza; ma la larghezza dell'intervallo di confidenza continua a dipendere soltanto dalla deviazione standard.