



OBIETTIVO

Obiettivo
dell'esercizio



DURATA

Durata
dell'esercizio



MODALITÀ DI SVOLGIMENTO

A casa o in classe, a
gruppi o singolarmente

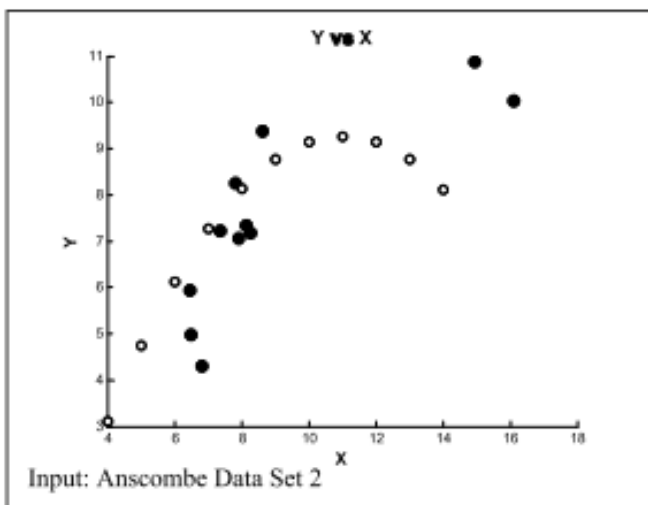
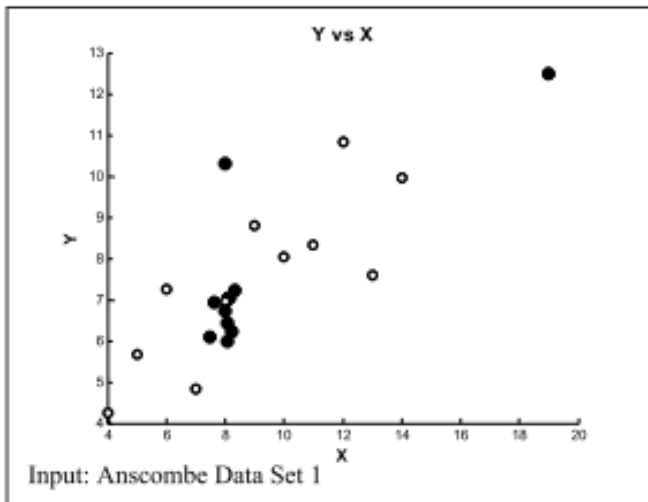
INFORMAZIONI

Dopo Anscombe, altri ricercatori hanno sviluppato metodi empirici per generare esempi simili. In particolare, nel **2007** due statistici della Northeastern University di Boston, **Sangit Chatterjee e Aykut Firat**, pubblicarono sulla stessa rivista sulla quale era uscito il Quartetto di Anscombe, *The American Statistician*, un metodo per sistematizzare la costruzione di esempi simili. L'articolo reca il titolo molto esplicativo **"Generating Data with Identical Statistics but Dissimilar Graphics: A Follow up to the Anscombe Dataset"**, e potete trovarlo [qui](#).

ISTRUZIONI

→ Step 1

Riportiamo nell'immagine due esempi di tale costruzione, tratti dall'articolo originale di Chatterjee e Firat, dove i pallini bianchi rappresentano il dataset in input e quelli neri il dataset in output:



La costruzione del dataset di output è realizzata mediante l'impiego di un algoritmo di ottimizzazione chiamato **algoritmo genetico**, che simula il comportamento di un gene soggetto a mutazioni attraverso generazioni successive, in ognuna delle quali solo le mutazioni più vantaggiose vengono conservate. In pratica, si parte da una soluzione molto approssimata e la si fa evolvere attraverso modifiche successive fino ad arrivare a una soluzione ottimale.

→ Step 2

Chatterjee e Firat forniscono anche alcune porzioni di codice MatLab per replicare la loro tecnica: chi di voi volesse cimentarsi con la programmazione potrebbe quindi provare, usando tale codice come guida, ad implementare il metodo in qualche linguaggio a voi familiare.



Le porzioni di codice MatLab dalle quali partire si trovano nell'articolo ["Generating Data with Identical Statistics but Dissimilar Graphics: A Follow up to the Anscombe Dataset"](#) di Chatterjee e Firat.