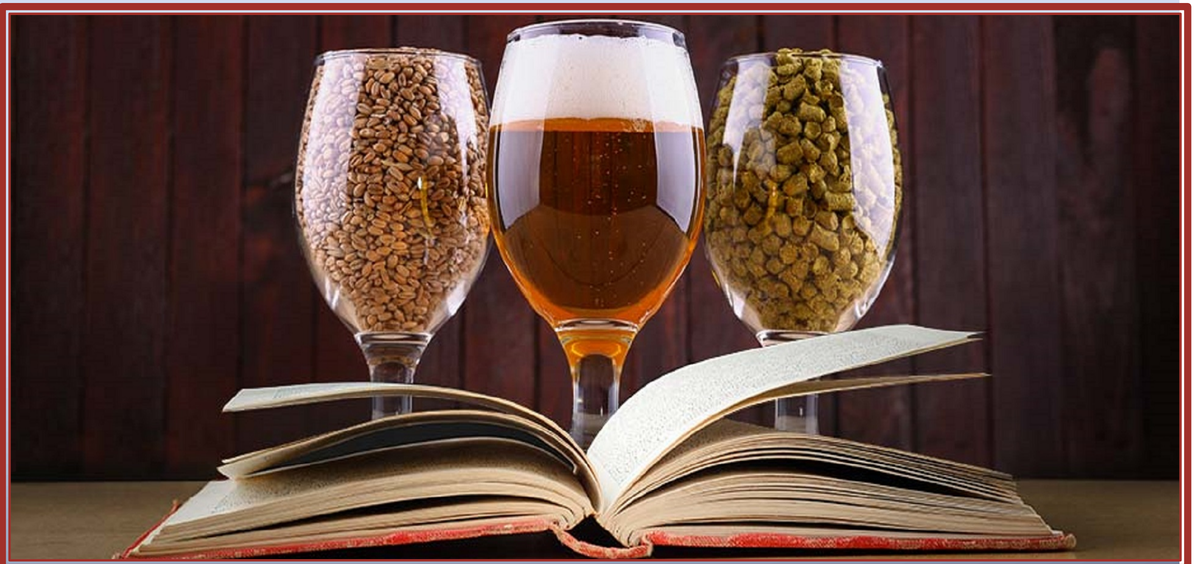
Sapienza Università di Roma
Facoltà di Scienze Matematiche Fisiche e Naturali
e
Liceo Scientifico Statale "A.Avogadro" Roma

PROGETTO Alternanza Scuola-Lavoro 2018 - 19

Parliamo di birra … (5)

Carlotta MAFFEI
Dipartimento di Matematica

Giuliana MASSOTTI
Liceo Scientifico Statale
"A.Avogadro" Roma



Nello scorso incontro abbiamo esplorato la possibilità di intraprendere un’attività di vendita di birra (artigianale o industriale) senza essere troppo esposti alla concorrenza.

Analizziamo ora come, servendosi ancora della matematica, sia possibile **caratterizzare la nostra offerta** per differenziarla da quella dei concorrenti.

CARATTERIZZARE L’OFFERTA

Ricordiamo che **capire il mercato,** cioè studiare i bisogni e i desideri del cliente, per produrre e fargli trovare, dove vuole, ciò che desidera, è molto importante.

Un'indagine molto interessante di carattere quantitativo può rivelarsi quella che studia le relazioni tra il consumo di birra e abitudini, interessi, stato sociale … dei consumatori.

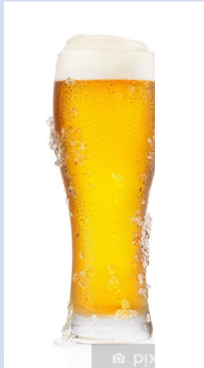
Dobbiamo iniziare la nostra analisi raccogliendo un po' di informazioni.

Supponiamo di voler indagare, ad esempio,

se i consumatori vanno più spesso (e volentieri) in un locale che è vicino alla loro abitazione

Intervistiamo 10 persone <u>scelte a caso</u> (nella realtà dovrebbero essere molte di più) nella zona prescelta e poniamo loro le seguenti domande (inchiesta di mercato):

“se una birreria è distante da casa vostra



(a) 150m.,
(b) 300m,
(c) 500m.,
(d) 700m. ,
(e) circa 1Km,“

andate in birreria

(a’) 1 volta alla settimana,
(b’) 2volte alla settimana,
(c’) 3 volte alla settimana,
(d’) 4 volte alla settimana,
(e’) 5 volte alla settimana.“

Le risposte sono le seguenti:

La prima persona intervistata sceglie la (a’) e la (a) (1 volta, 150m.),
La seconda persona sceglie la (a’) e la (b) (1 volta, 300m.)
La terza persona sceglie la (b’) e la (d), (2 volte, 700m.)
La quarta persona sceglie la (e’) e la (a), (5 volte, 150m.)
La quinta persona sceglie la (d’) e la (b), (4 volte, 300m.)
La sesta persona sceglie la (c’) e la (a), (3 volte, 150m.)
La settima persona sceglie la (b’) e la (b), (2 volte, 300m.)
La ottava persona sceglie la (e’) e la (b),   (5 volte, 300m.)
La nona persona sceglie la (b’) e la (c’), (2 volte , 500m.)
La decima persona sceglie la (a’) e la (e) (1 volta, 1000m.).

Vorremmo sapere se esiste una relazione tra il numero di volte che le persone vanno in birreria e la distanza della birreria dalla casa delle persone.

CHE SIGNIFICA "*Esiste una relazione (tra il numero delle volte che si va in birreria e la distanza)*?"

La relazione più semplice che si possa immaginare tra due insiemi di dati è quella che ci dice se

all'aumentare di un insieme di dati aumentano o diminuiscono anche i dati dell'insieme corrispondente.

Nel caso che stiamo studiando ci chiediamo, in particolare, se è vero che il numero di volte in cui si va in birreria diminuisce o aumenta all'aumentare della distanza della birreria

Per rispondere

rappresentiamo le risposte come punti di un piano cartesiano in cui sull'asse orizzontale è riportato il **n. di volte** che si va in birreria e su quello verticale c'è la **distanza** da casa:

A = (V,D) = (1, 150),

B = (1, 300) ,

C = (2, 700),

D = (5, 150),

E = (4, 300),

F = (3, 150),

G = (2, 300),

H = (5, 300),

I = (2, 500),

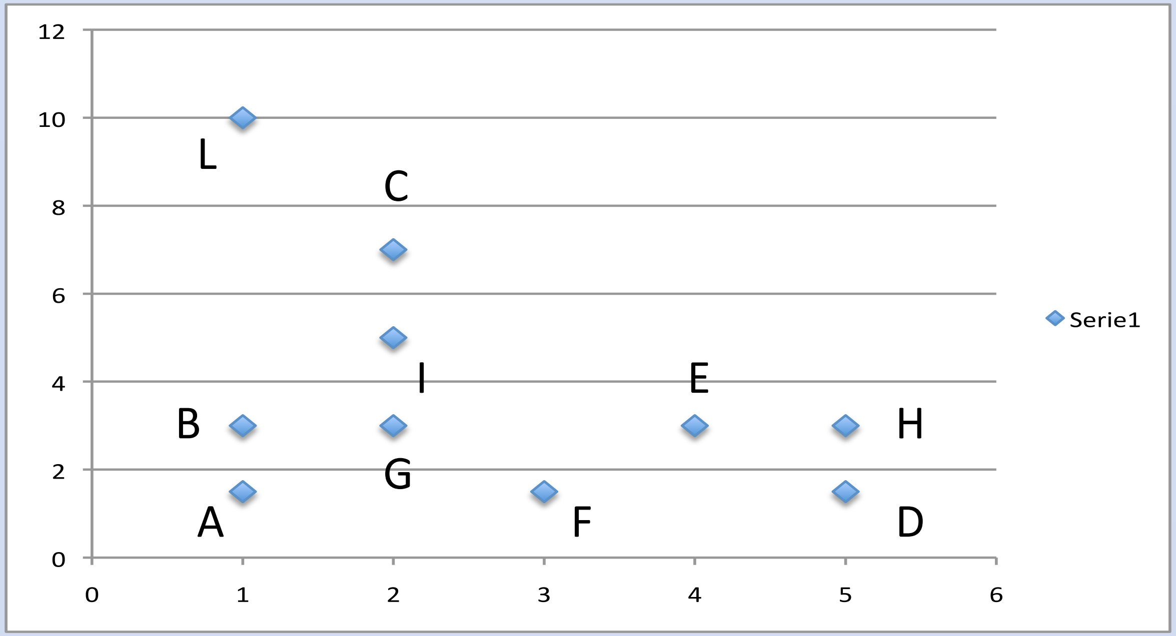
L = (1, 1000).

Diagramma di dispersione (o diffusione)

D

distanza
(in metri)

1 = 100 metri



V n. di volte

I punti sono “dispersi” nel piano

In queste interviste,

il numero medio di volte a settimana in si va in birreria è

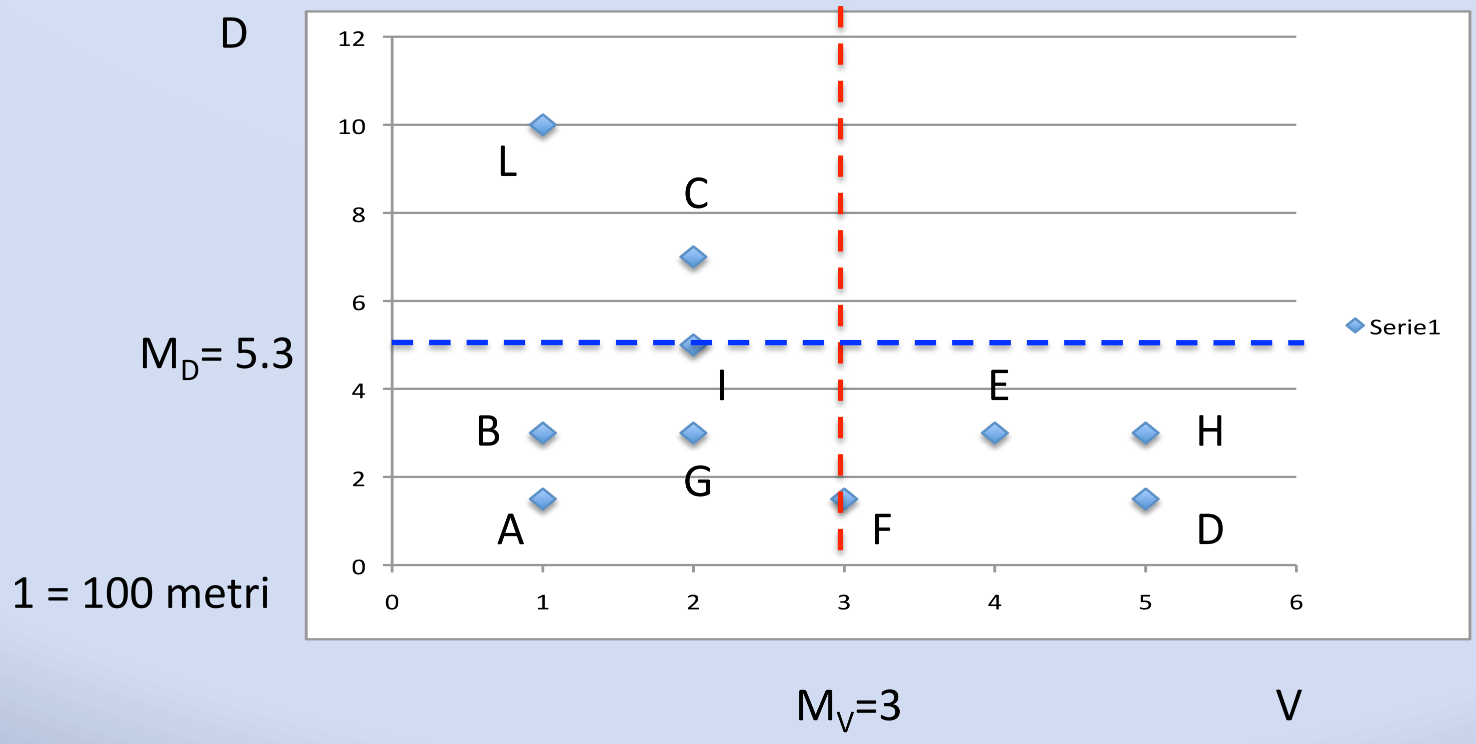
$\mathbf{M_V}$= (1 + 2 + 3 + 4 + 5)/5 = 3 volte,

La distanza media è

$\mathbf{M_D}$ = (150 + 300 + 500 + 700 + 1000)/5 = 530 m.

Rappresentiamo anche queste informazioni sul grafico

Diagramma di dispersione



I dati sono anche "dispersi" (cioè sono ad una certa distanza) rispetto alle medie.

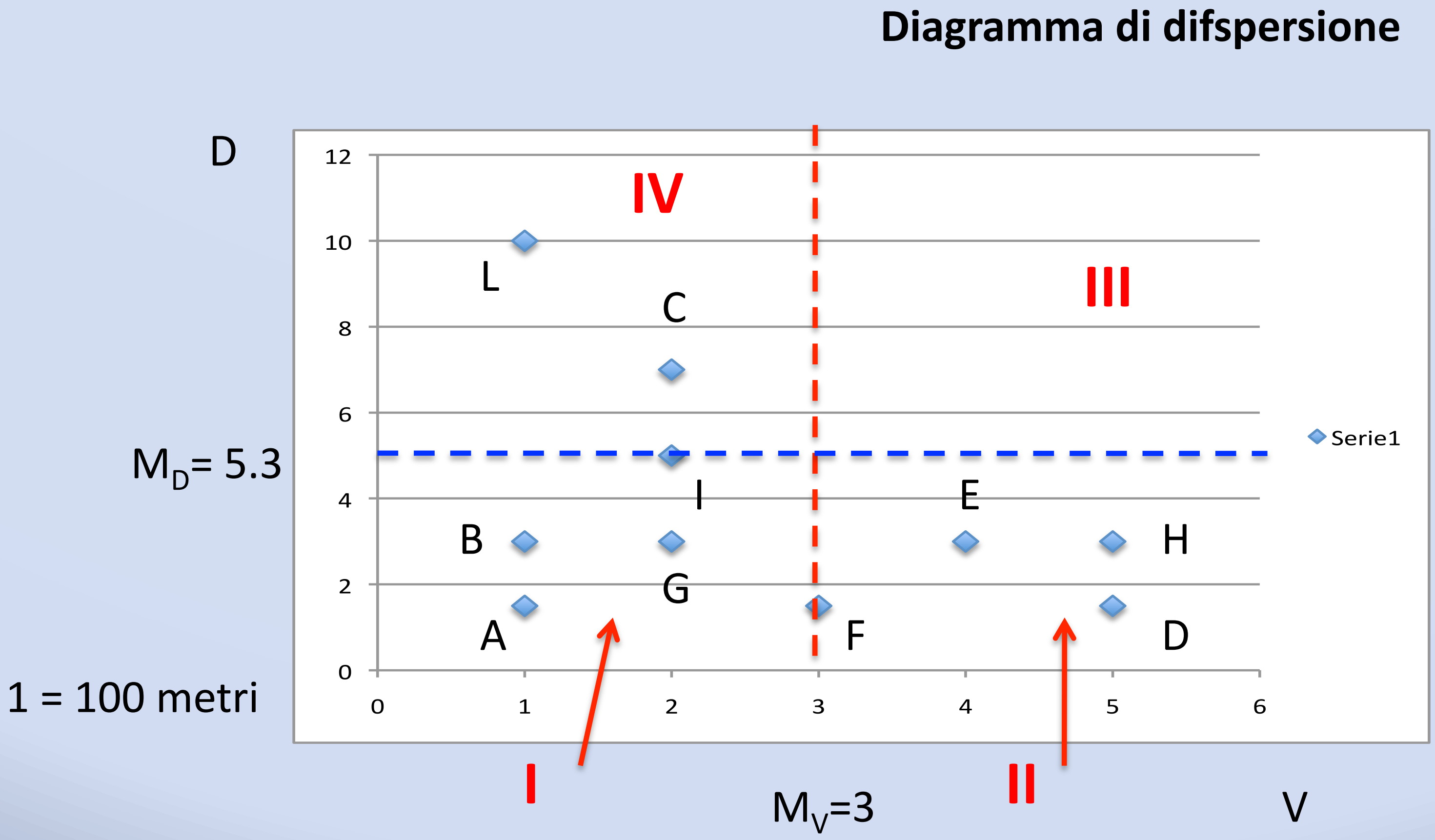
In particolare alcuni dati sono alla destra delle medie, altri sono alla sinistra, alcuni sono sopra, altri sono sotto.

Come tener conto della posizione dei punti rispetto alle rette delle medie?

Notiamo che se la prima coordinata di un punto (il n. di volte che si va in birreria) è più piccola del n. medio di volte che si va, la differenza $V_i – M_V$ è positiva, si ha il contrario se la prima coordinata di un punto è più grande.

Lo stesso accade per la seconda coordinata del punto.

Questo significa che se il punto ha tutte e due le coordinate minori delle medie il prodotto $(V_i – M_V)(D_i – M_V)$ è **positivo**: questo accade per i punti che si trovano nel primo dei IV quadranti in cui le rette delle medie dividono il piano



I: le due coordinate sono più piccole delle medie $(V_i - M_V)(D_i - M_V) > 0$

II: la prima coordinata è più grande della media, la seconda più piccola $(V_i - M_V)(D_i - M_V) < 0$

III: le due coordinate sono più grandi delle medie $(V_i - M_V)(D_i - M_V) > 0$

IV: la prima coordinata è più piccola della media, la seconda più grande $(V_i - M_V)(D_i - M_V) < 0$

Il segno dei prodotti $(V_i - M_V)(D_i - M_V)$ ci dà informazione sulla posizione dei punti del piano.

Come ricaviamo informazioni sull'eventuale relazione tra i due insiemi di dati?

Se la maggior parte dei punti si trova in I o in III (e quindi la maggior parte dei prodotti è positiva), vuol dire che all'aumentare della prima coordinata aumenta anche la seconda, se invece la maggior parte dei punti si trova in II o in IV (e quindi la maggior parte dei prodotti è negativa), vuol dire che all'aumentare della prima coordinata la seconda diminuisce.

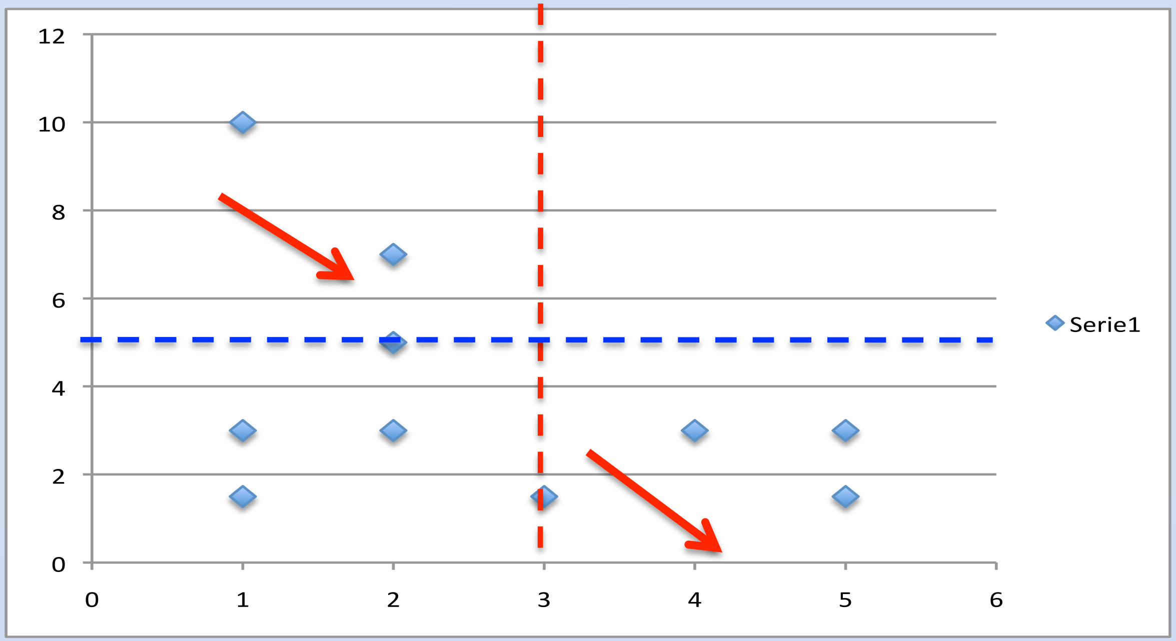
Nel caso che stiamo studiando cosa accade?

Diagramma di dispersione

D

M_D= 5.3

1 = 100 metri



M_V=3

V

La **maggior parte** dei punti (7 su 10) si trova nel settore II e IV: questo indica che all'aumentare del numero di volte la distanza diminuisce:

si può dire che prevediamo una dipendenza negativa tra le due variabili

Per avere una conferma quantitativa della conclusione precedente calcoliamo la **covarianza** tra i dati (co – variare = variare insieme).

DEFINIZIONE. Si definisce covarianza di N (N=1,2,3,…) coppie di dati (V,D) il seguente indice (misura)

$$C_{VD} = [(V_1 - M_V)(D_1 - M_D) + (V2 - M_V)(D_1 - M_D) + \dots + (V_N - M_V)(D_N - M_D)]/N$$

(la media dei prodotti delle distanze con segno dei dati dalla media).

- Se $C_{VD}>0$, i dati V e D sono **correlati positivamente** (cioè al crescere dei valori V, aumentano i valori D)
- se $C_{VD}<0$, i dati V e D sono **correlati negativamente** (cioè al crescere dei valori V, diminuiscono i valori D),
- se infine $C_{VD}=0$ i dati V e D sono **scorrelati**, cioè non vi è nessuna variazione comune tra di essi.

Calcoliamo la covarianza.

$V_1 - M_V = 1 - 3 = -2$	$D_1 - M_D = 150 - 530 = -380$	$(V_1 - M_V)(D_1 - M_D) = 760$
$V_2 - M_V = 1 - 3 = -2$	$D_2 - M_D = 300 - 530 = -230$	$(V_2 - M_V)(D_2 - M_D) = 460$
$V_3 - M_V = 2 - 3 = -1$	$D_3 - M_D = 700 - 530 = 170$	$(V_3 - M_V)(D_3 - M_D) = -170$
$V_4 - M_V = 5 - 3 = 2$	$D_4 - M_D = 150 - 530 = -380$	$(V_4 - M_V)(D_4 - M_D) = -760$
$V_5 - M_V = 4 - 3 = 1$	$D_5 - M_D = 300 - 530 = -230$	$(V_5 - M_V)(D_5 - M_D) = -230$
$V_6 - M_V = 3 - 3 = 0$	$D_6 - M_D = 150 - 530 = -380$	$(V_6 - M_V)(D_6 - M_D) = 0$
$V_7 - M_V = 2 - 3 = -1$	$D_7 - M_D = 300 - 530 = -230$	$(V_7 - M_V)(D_7 - M_D) = 230$
$V_8 - M_V = 5 - 3 = 2$	$D_8 - M_D = 300 - 530 = -230$	$(V_8 - M_V)(D_8 - M_D) = -460$
$V_9 - M_V = 2 - 3 = -1$	$D_9 - M_D = 500 - 530 = -30$	$(V_9 - M_V)(D_9 - M_D) = 30$
$V_{10} - M_V = 1 - 3 = -2$	$D_{10} - M_D = 1000 - 530 = 470$	$(V_{10} - M_V)(D0_1 - M_D) = -940$

$$C_{VD} = [\cancel{760} + \cancel{460} - 170 - \cancel{760} - \cancel{230} + \cancel{230} - \cancel{460} + 30 - 940]\backslash 10 = -1080/10 =$$
$$= -108 < 0$$

Questo numero fornisce la misura di quanto il numero di volte settimanale in cui si va in birreria **vari in dipendenza dalla distanza** del locale da casa.

Il risultato mostra che un assiduo frequentatore preferisce locali più vicini alla sua abitazione

Approfondimento di matematica

Date due collezioni di N dati $X = (x_1, x_2, \dots, x_N)$ e $Y = (y_1, y_2, \dots, y_N)$, dimostrare che valgono le seguenti proprietà:

1) se $Y=(y_1,y_2, \dots, y_N)$ **=** $(x_1,x_2, \dots, x_N)=X$ (**le due collezioni sono uguali**)

allora

la covarianza C_{XY} coincide con la varianza V della collezione X;

2) la covarianza C_{XY} è uguale alla covarianza C_{YX};

3) Se k è una costante qualunque la covarianza la covarianza $C_{Xk}=C_{Yk}=0$ (una collezione di dati e una costante qualsiasi sono sempre scorrelate)

Per esprimere l'intensità della relazione tra i dati di due collezioni si usa anche **il coefficiente di correlazione (o di Pearson)**

Definizione. Il **coefficiente di correlazione r_{XY}** di due collezioni di N dati $X=(x_1,x_2, \dots , x_N)$ e $Y=(y_1,y_2, \dots, y_N)$ è un indice numerico definito nel modo seguente

$$r_{XY} = C_{XY} /(\sigma_X\sigma_Y)$$

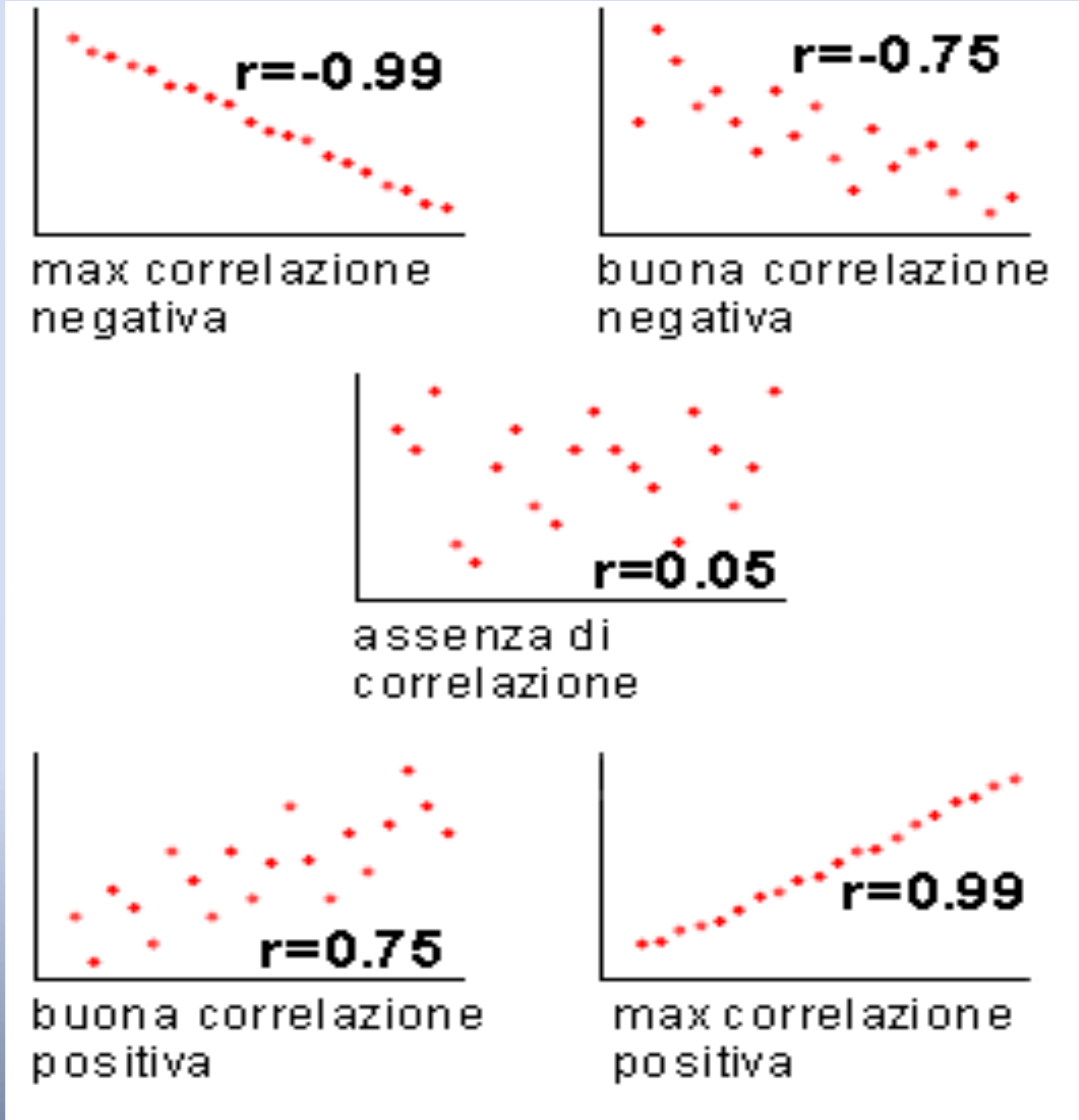
(il rapporto tra la covarianza e il prodotto delle deviazioni standard).

Si può dimostrare che il coefficiente di correlazione ha la seguente proprietà:

$$-1 \leq r_{XY} \leq 1,$$

Se $-1 \leq r_{XY} <0$, i dati sono **correlati negativamente** (la covarianza è negativa), se $0<r_{XY}\leq 1$, i dati sono **correlati positivamente** (la covarianza è positiva), se infine $r_{XY}=0$, i dati **non sono correlati** (la covarianza è nulla).
Si può anche dimostrare che

- Se $r_{XY} = 1$ i dati sono **allineati**, lungo una retta del primo e terzo quadrante,
- Se $r_{XY}= -1$ i dati sono **allineati**, lungo una retta del secondo e quarto quadrante.


r=-0.99
max correlazione negativa
r=-0.75
buona correlazione negativa
r=0.05
assenza di correlazione
r=0.75
buona correlazione positiva
r=0.99
max correlazione positiva

Osservazione 1.

Se $r_{XY} = 1$ (o se $r_{XY} = -1$) i dati sono allineati, cioè le coppie (x,y) sono punti di una retta. Questo vuol dire che la relazione tra x e y è del tipo y=ax+b: y è un multiplo di x a meno della costante b.
In questo caso si dice che tra i dati vi è una **relazione causale,** cioè una variazione di una delle variabili provoca una variazione conseguente e di tipo noto dell'altra.

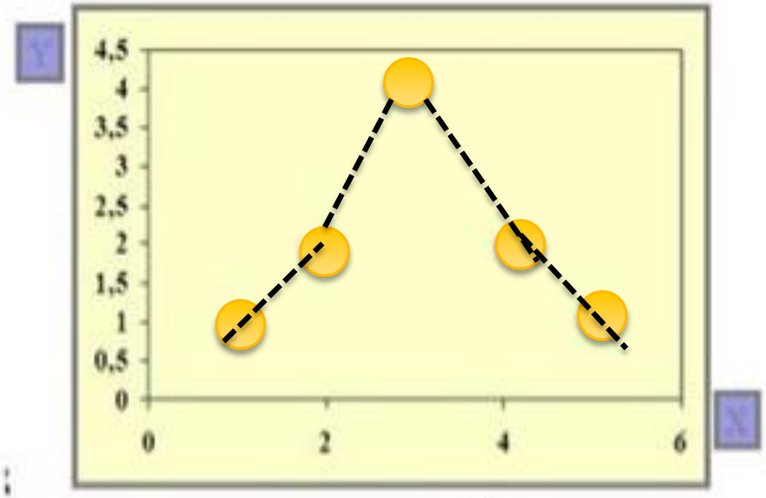
Se, come accade in generale, tra le due serie di dati non è identificabile una "causa ed un effetto", non si può più parlare di dipendenza tra le due serie di dati, ma si può misurare, tramite il coefficiente di correlazione, **l'intensità della associazione** tra i comportamenti delle due serie di dati.

Osservazione 2.

Se due serie di dati X e Y non sono correlate ($\rho_{XY} = 0$) non è detto che i dati siano tra loro scorrelati. Consideriamo infatti la situazione della figura

X	Y	$X-\mu_X$	$(X-\mu_X)^2$	$Y-\mu_Y$	$(Y-\mu_Y)^2$	$(X-\mu_X)(Y-\mu_Y)$
1	1	-2	4	-1	1	2
2	2	-1	1	0	0	0
3	4	0	0	2	4	0
4	2	1	1	0	0	0
5	1	2	4	-1	1	-2
			10		6	0

DEV(X) DEV(Y) COD(X,Y)



Media(X)=3; Media(Y)=4; Var(X)=2; Var(Y)=1,2; Cov(X,Y)=0; $\rho_{XY} = \frac{0}{\sqrt{2 \cdot 1,2}} = 0$

La covarianza (e quindi la correlazione) sono uguali a 0, ma è evidente che si può stabilire una relazione tra i dati tramite spezzate di rette (funzione definita "a tratti.)

La correlazione studia se la relazione tra i dati può avvicinarsi ad una relazione lineare (la più facile da stabilire).

Calcoliamo **il coefficiente di correlazione r_{VD}** dei dati V e D.

Abbiamo già calcolato la covarianza (C_{VD}= -108), calcoliamo le deviazioni standard
σ_V e $\sigma_{D.}$

Si ha

$\sigma_V = (V_V)^{1/2}=\{ [(V_1 - M_V)^2 +(V_2 - M_V)^2+ \ldots + (V_{10} - M_V)^2]/10\}^{1/2} = (24/10)^{1/2} \approx 1.55$

$\sigma_D = (V_D)^{1/2}=\{ [(D_1 - M_D)^2 +(D_2 - M_D)^2+ \ldots + (D_{10} - M_D)^2]/10\}^{1/2} =$
$= (895500/10)^{½} \approx 299.25$

$\sigma_V\sigma_D \approx 463.84$ e quindi $r_{VD} = C_{VD}/(\sigma_V\sigma_D) \approx - 0.65<0$:

come ci aspettavamo, tra le due serie di dati vi è una **correlazione negativa moderata**, che conferma, nel caso in studio, che più il locale è vicino casa, tanto più spesso lo si potrebbe frequentare.

Questa conclusione dice che, nel caso in studio, sarebbe opportuno aprire l'attività in vicinanza di un posto abitato, piuttosto che in un posto magari bello (giardino, …) ma lontano dalle abitazioni.

La stessa analisi si può ripetere per studiare l'eventuale relazione tra il numero di volte che le persone vanno in birreria e i prezzi; oppure tra il numero di volte che le persone vanno in birreria e l'eventuale cibo che viene servito nel locale.

Approfondimento di statistica

Supponiamo che vogliate reclamizzare l'apertura del vostro locale con messaggi pubblicitari da mandare in onda per una settimana su varie emittenti radio. Per capire se questa iniziativa può avere senso un esperto di marketing vi sottopone i risultati di un precedente sondaggio molto simile

Emittente	Messaggi al giorno	Aumento n. clienti al giorno
Fox	4	150
Roma	2	80
Aria libera	5	210
Gaia	6	240
Jumper	3	170

Quali informazioni si possono dedurre da questa tabella?

Per rispondere,

(a) rappresentare i dati come punti in un piano cartesiano scegliendo come unità di misura per il numero dei clienti del locale 10 = 1.

(b) Interpretare il diagramma di diffusione che si ottiene e quantificare l'intensità dell'eventuale associazione tra la frequenza giornaliera dei messaggi pubblicitari e il numero di nuovi clienti calcolando covarianza e coefficiente di correlazione.

Conviene investire soldi in questa forma di pubblicità?

L’interessante storia dell’idea di correlazione e l’eugenetica

Il concetto di correlazione è legato alla ricerca compiuta da uno studioso inglese, cugino di Charles Darwin: **Francis Galton** (1822 – 1911).

La famiglia Galton era benestante e Francis fu educato in ottime scuole. Studiò inizialmente Medicina ma dal 1840 al 1844 si dedicò agli studi di matematica, che seguì a Cambridge.
A partire degli anni ‘60 dell’800 i suoi interessi di ricerca s'indirizzarono sempre più verso il problema dell'eredità dei caratteri fisici e mentali.
L'evoluzionismo di *On the Origin of Species,* scritto dal cugino e pubblicato nel 1859 e gli studi statistici sui caratteri misurabili degli esseri umani, esercitarono enorme influenza sui suoi studi.

Le indagini di Galton sull'eredità hanno prodotto molte opere notevoli: *Hereditary Genius* (1869), *English Men of Science* (1874), *Inquiries into Human Faculty and Its Development* (1883) e *Natural Inheritance* (1889).

In quest'ultima opera, nell'ultimo capitolo, Galton "**immaginò lo sviluppo d'una <u>teoria quantitativa</u> predittiva della discendenza** che collegasse la generazione dei genitori con quella della prole".

In anni a cavallo tra il 1870 e il 1880 Galton si dedicò pazientemente a raccogliere informazioni sulla statura perché era "*facile da misurare, relativamente costante durante la vita adulta*". Introdusse il coefficiente di correlazione per studiare l'andamento di questa grandezza in relazione ad altre, come la lunghezza delle braccia, e per trovare quanto fossero frequenti i casi lontani dalla media.

PERCHE' GALTON ERA INTERESSATO AI CASI LONTANI DALLA   MEDIA?

Le indagini statistiche di Galton avevano come obiettivo quello di provare quanto fossero "**stabili attraverso le generazioni**" certe caratteristiche degli esseri umani .

Per l'altezza in particolare (ma anche per tutti gli altri caratteri umani), la statistica mostrava che nel tempo questa si stabilizzava intorno al valore medio, che non sempre caratterizzava l'eccellenza.

Se si voleva quindi conservare qualche caratteristica ritenuta "migliore" ad esempio un'alta statura, secondo Galton, si sarebbero dovuti prendere provvedimenti, come una procreazione selettiva, indirizzati a mantenere questa caratteristica.

Secondo Galton, l'evoluzione predetta da Darwin, andava "presa per mano", forzata, per attuare un miglioramento della specie umana. Per raggiungere questo obiettivo bisognava incoraggiare l'eccellenza sia intellettuale che sociale e favorire la procreazione di portatori di essa, visto che "*il genio tende ad essere ereditario*".

L'idea di un progetto moderno di miglioramento della popolazione umana attraverso una comprensione statistica dell'ereditarietà, incoraggiata da un buon "allevamento", venne detta "**eugenetica**" (dal greco *della buona nascita*).

La prima teoria eugenetica promuoveva esplicitamente un modello di uomo che corrispondeva per l'appunto al gruppo sociale da cui proveniva Galton: l'élite della società britannica rappresentata dai liberi professionisti, dalle antiche famiglie dell'aristocrazia terriera e dagli scienziati.
Le nuove ricchezze, costruite per l'industria e il commercio, non trovano invece alcun favore ai suoi occhi.

Politicamente l'eugenetica galtoniana appare come una teoria conservatrice che ha lo scopo primario di proteggere un gruppo sociale ritenuto migliore, contro la minaccia proveniente dagli strati più bassi della popolazione.

Sotto l'**apparenza** di natura scientifica la teoria ricercava giustificazioni per la preservazione ed il mantenimento dell'ordine sociale vigente, che esige limiti severi per le unioni tra individui di contesti sociali differenti.

EUGENETICA E RAZZISMO

Il razzismo e l'eugenetica spesso si mescolano negli argomenti degli eugenetici conservatori. Galton, come molti dei suoi contemporanei, assunse il fatto di essere inglese come "condizione razziale" privilegiata; per la sua teoria, risalente ai primi anni cinquanta, non può però essere stato in alcun modo influenzato dalle idee di Darwin, pubblicate per la prima volta nel 1859.

All'inizio del XX secolo la preoccupazione nei riguardi del "deterioramento nazionale" venne rafforzata dalla creazione di strumenti statistici per la misurazione dei giovani soldati di leva. Sulla base di queste cifre si ritrova regolarmente una "degenerazione fisica ed intellettuale" della popolazione, con un'inquietudine rivolta particolarmente alle differenze dei tassi di fecondità tra i popoli nordici e i recenti migranti dell'epoca, provenienti dall'Europa orientale.
La paura nei confronti della fecondità delle classi popolari si accompagna a preoccupazioni circa gl'immigrati cattolici irlandesi, ebrei, polacchi e russi, alimentando un antisemitismo sempre latente.

La storia ci dice che il regime nazista consumò tragicamente il matrimonio tra razzismo ed eugenetica, colpendo con le sue leggi di sterilizzazione innanzitutto i mulatti nati durante l'Occupazione della Ruhr da parte delle truppe coloniali francesi nel 1923, un fatto denunciato come "*vergogna negra*".

Con l'avvento del nazionalsocialismo e con l'applicazione metodica del programma per l'eliminazione delle "razze inferiori",
dal 1933 al 1945 (fine della II Guerra Mondiale)
l' "*igiene razziale*" sarà la giustificazione lo sterminio fisico di milioni di ebrei, zingari, omosessuali e avversari politici.

Approfondimento di storia

Partendo dall’antica Grecia, ricostruire le radici dell’eugenetica.

L’eugenetica in Europa e negli Stati Uniti.

Degenerazioni dell’eugenetica

Approfondimento di genetica

La genetica moderna ha dimostrato che il concetto di differenti razze umane non ha fondamento scientifico : i risultati di L. L. Cavalli Sforza

Il nostro percorso attraverso birre, lieviti, matematica storia, statistica, chimica eccetera finisce qui.

Speriamo che quello che abbiamo fatto insieme possa esservi utile, non tanto per intraprendere il mestiere di birraio, quanto per farvi riflettere sul fatto che la conoscenza “ad ampio raggio” (interdisciplinare) è indispensabile per avere successo nella vita e, soprattutto, per farvi vivere meglio, più consapevoli, più liberi e più felici.

Come ha detto Steve Jobs, il fondatore della Apple, parlando ad una platea di studenti universitari
“ragazzi, il vostro tempo è limitato, quindi non sprecatelo vivendo la vita di qualcun altro. Siate affamati [di conoscenza], siate folli, perché solo coloro che sono abbastanza folli da pensare di poter cambiare il mondo lo cambiano davvero.”

GRAZIE DELLA VOSTRA ATTENZIONE!