



Matematica in laboratorio

1. LO STUDIO DELLA CONCENTRAZIONE

Una delle applicazioni più significative dello studio della concentrazione è relativa alla distribuzione del reddito. Supponiamo, ad esempio, che il reddito di 382 lavoratori dipendenti (anche stagionali o saltuari) di una certa zona possa essere rappresentato dalla tabella seguente:

| Classi di reddito (in migliaia di €) | Frequenza |
|---|------------|
| [0-5) | 6 |
| [5-10) | 28 |
| [10-15) | 34 |
| [15-20) | 98 |
| [20-25) | 102 |
| [25-30) | 66 |
| [30-40) | 21 |
| [40-50) | 18 |
| [50-60] | 9 |
| TOTALE | 382 |

Valutiamo la concentrazione di questa distribuzione con Excel; segui la procedura indicata di seguito.

Nelle colonne A e B devi inserire rispettivamente l'estremo sinistro e l'estremo destro di ciascuna delle classi di reddito e nella colonna C devi inserire le frequenze.

Nella cella C11 calcoliamo poi la somma delle frequenze usando la funzione di somma automatica individuata dall'i-

cona .

Nella colonna D calcoliamo i valori centrali delle classi:

D2: $= (A2+B2) / 2$ Copia la formula nelle celle da D3 a D10.

Nella colonna E calcoliamo l'ammontare complessivo dei redditi per ciascuna classe:

E2: $= D2 * C2$

Copia poi la formula nelle celle da E3 a E10 e in E11 calcola il totale di colonna.

Nella colonna F calcoliamo le frequenze relative percentuali dei lavoratori:

F2: $= C2 * 100 / \$C\11

Copia la formula per la parte di colonna interessata e calcola la somma dei suoi valori in F11.

Nella colonna G calcoliamo adesso le frequenze relative percentuali dei redditi:

G2: $= E2 * 100 / \$E\11

Copia la formula e calcola poi la somma dei valori della colonna in G11.

Nelle colonne H ed I calcoliamo le frequenze cumulate rispettivamente dei lavoratori e dei redditi:

H2: $= F2$

H3: $= H2+F3$ Copia adesso la formula della cella H3 da H4 a H10.

Per costruire la colonna I puoi copiare la colonna H nelle celle da I2 a I10.

La tabella dei dati che ci necessitano è ora completa; procediamo al calcolo dell'area e del rapporto di concentrazione (vedi la figura a lato).

A13: AREA T1

A14: AREA T2

Seleziona le due celle A13 e A14 e trascina la selezione fino alla cella A21; in questo modo hai generato la serie dei trapezi di cui dobbiamo calcolare l'area. Per eseguire tale calcolo, inserisci nelle celle indicate le seguenti formule:

C13: $= l2 * H2 / 2$ area del triangolo T1

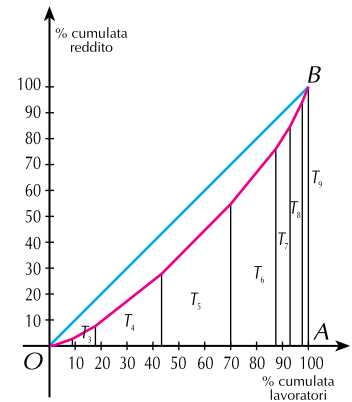
C14: $= (l2+l3) * (H3-H2) / 2$ area del trapezio T2

Copia adesso la formula di C14 da C15 a C21.

In H16 e H17 possiamo finalmente calcolare l'area ed il rapporto di concentrazione:

H16: $= 5000-SOMMA(C13:C21)$

H17: $= H16 / 5000$



| | A | B | C | D | E | F | G | H | I |
|----|---------|----|----------|-----------|----------------------------|-------------|-------------|--------------|--------------|
| 1 | CLASSI | | FREQ. | V. CENTR. | REDDITI | F.R. LAV. | F.R. RED. | F. CUM. LAV. | F. CUM. RED. |
| 2 | 0 | 5 | 6 | 2,5 | 15 | 1,570680628 | 0,176159718 | 1,570680628 | 0,176159718 |
| 3 | 5 | 10 | 28 | 7,5 | 210 | 7,329842932 | 2,466236054 | 8,90052356 | 2,642395772 |
| 4 | 10 | 15 | 34 | 12,5 | 425 | 8,90052356 | 4,991192014 | 17,80104712 | 7,633587786 |
| 5 | 15 | 20 | 98 | 17,5 | 1715 | 25,65445026 | 20,14092777 | 43,45549738 | 27,77451556 |
| 6 | 20 | 25 | 102 | 22,5 | 2295 | 26,70157068 | 26,95243688 | 70,15706806 | 54,72695244 |
| 7 | 25 | 30 | 66 | 27,5 | 1815 | 17,27748691 | 21,3153259 | 87,43455497 | 76,04227833 |
| 8 | 30 | 40 | 21 | 35 | 735 | 5,497382199 | 8,631826189 | 92,93193717 | 84,67410452 |
| 9 | 40 | 50 | 18 | 45 | 810 | 4,712041885 | 9,51262478 | 97,64397906 | 94,1867293 |
| 10 | 50 | 60 | 9 | 55 | 495 | 2,356020942 | 5,813270699 | 100 | 100 |
| 11 | TOTALI | | 382 | | 8515 | 100 | 100 | | |
| 12 | | | | | | | | | |
| 13 | AREA T1 | | 0,138 | | | | | | |
| 14 | AREA T2 | | 10,330 | | | | | | |
| 15 | AREA T3 | | 45,731 | | | | | | |
| 16 | AREA T4 | | 454,188 | | AREA CONCENTRAZIONE | | | 1166,559 | |
| 17 | AREA T5 | | 1101,459 | | RAPPORTO DI CONCENTRAZIONE | | | 0,23331171 | |
| 18 | AREA T6 | | 1129,682 | | | | | | |
| 19 | AREA T7 | | 441,760 | | | | | | |
| 20 | AREA T8 | | 421,400 | | | | | | |
| 21 | AREA T9 | | 228,754 | | | | | | |

2. LA RETTA DEI MINIMI QUADRATI CON EXCEL

I candidati ad un posto di tecnico di laboratorio in una Azienda Sanitaria vengono sottoposti ad un test di competenza; i risultati sono indicati nel foglio di lavoro di Excel che segue dove:

- nella prima colonna sono indicati gli anni di esperienza su un certo tipo di macchinario
- nella seconda il punteggio ottenuto nel test.

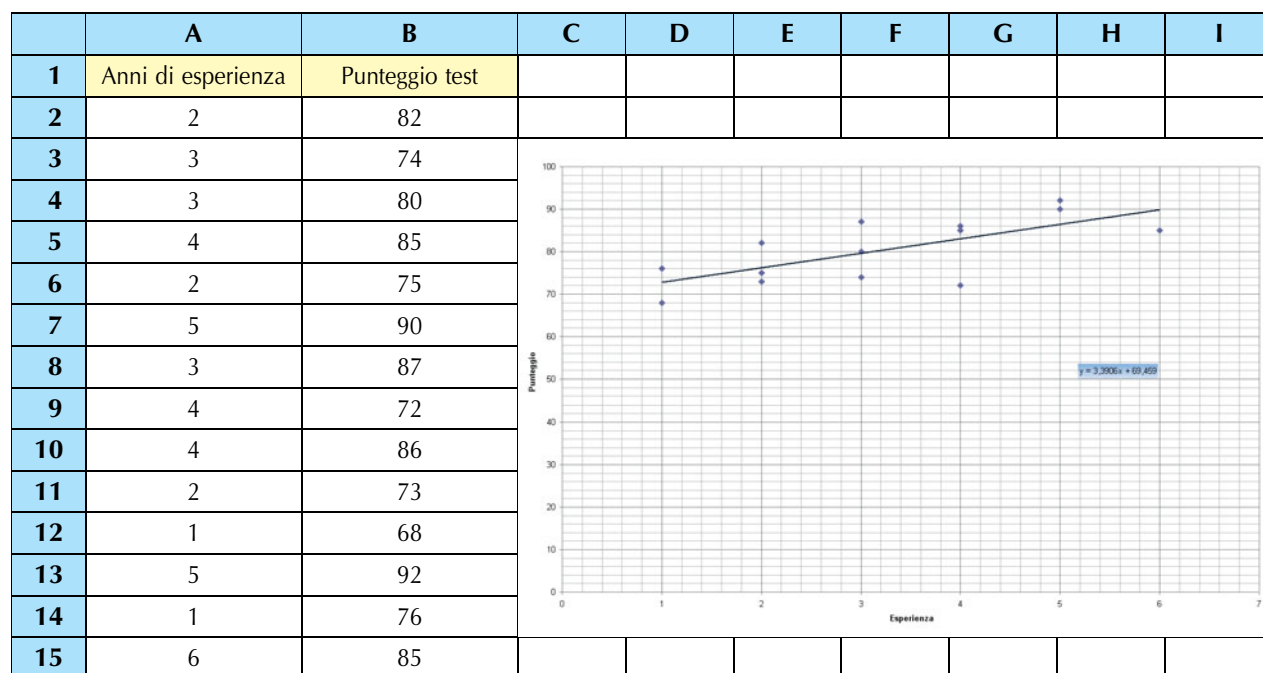
Vogliamo rappresentare questi dati e costruire la funzione di interpolazione lineare.
Per costruire il grafico:

- selezioniamo le due colonne dei dati
- dal menu **Inserisci** selezioniamo il *grafico a dispersione*

La costruzione della retta interpolante si può fare in modo molto semplice, ma la procedura cambia leggermente a seconda della versione di Excel usata.

In quella più recente basta scegliere tra le icone disponibili nel menu quella dove si trova disegnata la linea di tendenza; in versioni meno aggiornate si deve cliccare con il tasto destro del mouse su uno dei punti rappresentati e selezionare il comando *Aggiungi linea di tendenza* scegliendo poi tra quelle proposte quella lineare.

Si può poi visualizzare l'equazione della retta con un doppio clic sulla linea e mettendo il segno di spunta sulla voce *Visualizza l'equazione sul grafico* che si trova nella parte inferiore della finestra di dialogo.



3. LA DIPENDENZA STATISTICA CON EXCEL

Excel ha delle funzioni predefinite anche per il calcolo della covarianza, dell'indice di correlazione lineare e dei parametri della retta di regressione.

■ CORRELAZIONE (matrice1; matrice2)

È la funzione che determina il coefficiente di correlazione fra due insiemi di dati, che sono specificati nei parametri "matrice1" e "matrice2"; ciascun parametro è di solito individuato mediante un intervallo di celle. Ad esempio, CORRELAZIONE (A1:A10; B1:B10) considera come primo insieme di dati l'intervallo A1:A10, come secondo insieme di dati l'intervallo B1:B10.

■ COVARIANZA (matrice1; matrice2)

È la funzione che calcola la covarianza di due insiemi di dati; i parametri "matrice1" e "matrice2" hanno lo stesso significato di quelli della funzione precedente.

■ REG.LIN (ynota; xnota; cost; stat)

Restituisce i parametri della retta di regressione. Il parametro "ynota" è l'elenco dei valori della variabile Y, il parametro "xnota" è l'elenco dei valori della variabile X; nel nostro caso, gli elenchi saranno inseriti ciascuno in una colonna (o in una riga). Il parametro "cost" è un valore logico che se è FALSO fa assumere il valore 0 al parametro b , cioè al termine noto della retta (in sostanza viene richiesto che essa passi per l'origine degli assi); se è VERO oppure se

viene omesso fa calcolare b con la formula nota (vedi pagina 469). Il parametro "stat" è anch'esso un valore logico che se è VERO fa in modo che vengano restituiti anche alcuni parametri della regressione (errore standard, gradi di libertà e così via); se è FALSO o se viene omesso non fa calcolare tali valori; nel nostro caso non useremo questo parametro perché non abbiamo studiato nella parte teorica il loro significato. Invertendo i primi due parametri vengono calcolati i valori della retta di regressione di X su Y .

■ PENDENZA (ynota; xnota)

Determina il coefficiente angolare della retta di regressione di Y su X . Invertendo i parametri viene calcolato il coefficiente angolare della retta di regressione di X su Y .

■ INTERCETTA (ynota; xnota)

Determina l'ordinata all'origine della retta di regressione di Y su X . Invertendo i parametri viene calcolata l'ordinata all'origine della retta di regressione di X su Y .

Ti presentiamo ora un esempio di studio di dipendenza fra due variabili statistiche in cui impareremo a calcolare i parametri di correlazione e di regressione lineare. La produzione di un certo tipo di deodorante in stick viene fatta ad una temperatura che può variare fra 80°C e 84°C; misurando la durezza del materiale prodotto si è notato che essa non è costante e ci si chiede, di conseguenza, se esiste una dipendenza fra la temperatura e la durezza e, in caso di risposta affermativa, se si tratta di una dipendenza lineare.

Per rispondere al quesito si rilevano le temperature in fase di lavorazione e la durezza su un campione di 36 pezzi prodotti; i valori di tali rilevazioni vengono poi riportati in una tabella.

Vediamo come dobbiamo procedere all'analisi dei dati con l'aiuto di Excel.

Innanzitutto devi inserire i dati come ti è indicato nella figura di pagina seguente (colonne A, B, C); fatto ciò, a fianco della tabella dei dati, costruisci il grafico della distribuzione scegliendo il tipo a dispersione per avere la nube di punti (i dati sono quelli delle colonne B e C) e completa con la costruzione della retta di interpolazione.

Calcoliamo adesso il coefficiente di correlazione:

C26: =CORRELAZIONE(B2:B23;C2:C23)

dove il gruppo di celle B2:B23 costituisce la zona dei dati X ed il gruppo C2:C23 è la zona dei dati Y . Il valore trovato indica che vi è correlazione positiva fra le due variabili statistiche, ma che tale dipendenza non è molto forte.

Calcoliamo ora i parametri della regressione lineare di Y su X . Osserviamo che questa funzione ha almeno due valori in uscita, non sarà quindi possibile seguire la procedura consueta di immissione di una formula, perché dovremo indicare quali sono le zone del foglio che dovranno contenere i risultati; tali zone inoltre non possono trovarsi sparpagliate nel foglio di lavoro, ma devono costituire una matrice rettangolare.

Procedi allora in questo modo.

■ Dovendo calcolare i valori a e b della retta di regressione abbiamo bisogno di due celle contigue in cui inserire i risultati; seleziona dunque le celle C27 e D27.

■ Nella barra della formula digita ora la funzione =REGR.LIN(C2 : C23; B2 : B23)

■ Per confermare devi usare i tasti **CTRL-SHIFT** e poi **INVIO** (attenzione: non usare solo Invio).

Con questa procedura abbiamo inserito una "formula matrice", cioè una formula i cui risultati sono restituiti sotto forma di matrice.

Con lo stesso procedimento puoi ora calcolare i parametri della retta di regressione di X su Y (ricorda che nella formula devi invertire i dati X con quelli Y).

Un'altra funzione statistica di utilità è quella che consente di prevedere un valore della variabile y per un valore di x non osservato, sulla base dei parametri della retta di regressione. La sua sintassi è la seguente

PREVISIONE(x; ynota; xnota)

dove x è il valore per il quale si vuole calcolare y .

Per esempio, se si vuole conoscere la durezza prevista dello stick per una temperatura di 82,5° si deve impostare la funzione

=PREVISIONE(82,5; C2:C23; B2:B23)

Il valore restituito è 34,73064; tale valore è stato calcolato mediante la retta di regressione $y = 1,066952x - 53,2929$.

| | A | B | C | D | E | F | G | H | I | J |
|----|---------------------|-------------|---------|----------|---|---|---|---|---|---|
| 1 | n. pezzo | temperatura | durezza | | | | | | | |
| 2 | 1 | 81,3 | 31,5 | | | | | | | |
| 3 | 2 | 80,6 | 32,5 | | | | | | | |
| 4 | 3 | 81,4 | 35,6 | | | | | | | |
| 5 | 4 | 83,4 | 37,5 | | | | | | | |
| 6 | 5 | 84 | 34,6 | | | | | | | |
| 7 | 6 | 83,8 | 37,2 | | | | | | | |
| 8 | 7 | 80,7 | 31,4 | | | | | | | |
| 9 | 8 | 81,5 | 35,8 | | | | | | | |
| 10 | 9 | 82,6 | 35,9 | | | | | | | |
| 11 | 10 | 83 | 36,6 | | | | | | | |
| 12 | 11 | 81,1 | 32,2 | | | | | | | |
| 13 | 12 | 82 | 31,9 | | | | | | | |
| 14 | 13 | 81,9 | 33,5 | | | | | | | |
| 15 | 14 | 83,2 | 38 | | | | | | | |
| 16 | 15 | 83,6 | 34,3 | | | | | | | |
| 17 | 16 | 83 | 31,5 | | | | | | | |
| 18 | 17 | 82,5 | 34,6 | | | | | | | |
| 19 | 18 | 81,9 | 37,9 | | | | | | | |
| 20 | 19 | 81,3 | 35,1 | | | | | | | |
| 21 | 20 | 81 | 35,4 | | | | | | | |
| 22 | 21 | 80,6 | 33,3 | | | | | | | |
| 23 | 22 | 82,3 | 34,6 | | | | | | | |
| 24 | | | | | | | | | | |
| 25 | | | | | | | | | | |
| 26 | Coeff. correlazione | | 0,4666 | | | | | | | |
| 27 | Regressione Y su X | | 0,9243 | -41,3207 | | | | | | |
| 29 | Regressione X su Y | | 0,2355 | 73,9775 | | | | | | |

