

**Esercizi Foglio 8 - Statistica descrittiva - Distribuzioni a due caratteri**

**a) Valori sintetici di una distribuzione di dati: indici di posizione e indici di dispersione**

1) In un laboratorio ci sono 6 ratti, con code di lunghezza 11 cm, 8 cm, 7 cm, 7 cm, 12 cm, 9 cm. Calcola media, moda, mediana, varianza e deviazione standard della coda dei ratti.

(Soluz.:  $\bar{x} = 9 \text{ cm}$ ,  $\sigma^2 \simeq 3.67 \text{ cm}^2$ ,  $\sigma \simeq 1.91 \text{ cm}$ , mediana=8.5 cm, moda = 7 cm.)

2) Costruisci la tabella delle frequenze assolute, relative, percentuali, cumulate dei seguenti dati, che rappresentano i voti di maturità (in centesimi) di una classe di 20 studenti:

60 75 82 60 80 80 92 92 80 85 100 100 100 80 82 60 97 97 75 60

Calcola, inoltre:

a) media, moda, mediana e scarto interquartile (e disegna il box-plot dei dati)

b) 20-esimo e 76-esimo percentile.

(Soluzione:

voti	frequenze	frequenze cumulate
60	4	4
75	2	6
80	4	10
82	2	12
85	1	13
92	2	15
97	2	17
100	3	20

a)  $\bar{x} = 81.85$ , distribuzione bimodale con mode 60 e 80; mediana=81; per il calcolo del primo quartile, che corrisponde al 25-esimo percentile, calcoliamo  $np = 20 \times 0.25 = 5$ , da cui si ricava che  $q_1$  è la media tra il quinto e sesto valore, ovvero  $q_1 = 75$ ; analogamente, per il calcolo del terzo quartile, poiché  $np = 20 \times 0.75 = 15$ , si ha che  $q_3$  è la media tra il quindicesimo e sedicesimo valore, ovvero  $q_3 = 94.5$ ; scarto interquartile= 19.5; b) poiché  $np = 20 \times 0.20 = 4$ , il 20-esimo percentile è la media tra il quarto e quinto valore, cioè 67.5; essendo  $np = 20 \times 0.76 = 15.2$  il 76-esimo percentile è il sedicesimo dato, cioè 97.)

3) A pochi mesi dalla laurea, uno studente ha sul libretto i seguenti voti, con relativi crediti formativi universitari (CFU):

Voto	23	19	25	30	28	29	30	27	28	24	28	30	28
CFU	9	9	9	12	6	9	6	6	6	9	12	9	6

Qual è la media dei suoi voti? Per concludere il suo percorso di studi, lo studente deve sostenere un ultimo esame da 12 CFU. Se vuole presentarsi in seduta di laurea con almeno 27 di media, qual è il minimo voto che può accettare all'ultimo esame?

(Soluz: media= 26.7, il minimo voto che lo studente può accettare è 29)

4) I policlorobifenili (PCB) sono composti chimici ecotossici spesso presenti nel grasso corporeo. In uno studio sono state esaminate 14 persone e sono state rilevate le seguenti concentrazioni di PCB

(esprese in ng per g di lipidi):

520 560 720 900 930 1300 1700 1700 1800 1800 200 2500 3000 3200.

Calcola:

- a) l'intervallo di variazione, la media, la varianza e la deviazione standard campionarie;
- b) la mediana, il primo e terzo quartile, lo scarto interquartile;
- c) il 20-esimo e il 60-esimo percentile.

5) Si considerino i seguenti dati che rappresentano gli stipendi annuali iniziali (in migliaia di dollari) guadagnati da 42 ingegneri neolaureati:

Stipendi	27	28	29	30	31	32	34	36	37	40
Frequenze	4	1	3	5	8	10	5	2	3	1

Calcola lo stipendio medio e la deviazione standard, la mediana e lo scarto interquartile.

(Si ricordino le seguenti formule per calcolare media, varianza e deviazione standard nel caso di dati forniti con le rispettive frequenze:

$$\bar{x} = \frac{\sum_{i=1}^n f_i \cdot x_i}{\sum_{i=1}^n f_i}, \quad \sigma^2 = \frac{\sum_{i=1}^n f_i \cdot (x_i - \bar{x})^2}{\sum_{i=1}^n f_i}, \quad \sigma = \sqrt{\frac{\sum_{i=1}^n f_i \cdot (x_i - \bar{x})^2}{\sum_{i=1}^n f_i}},$$

dove  $f_i$  indica la frequenza relativa al dato  $x_i$ , per  $i = 1, \dots, n$ .)

## b) Distribuzioni a due caratteri e regressione lineare

Richiamo formule:

La retta di regressione (o retta dei minimi quadrati) della variabile  $Y$  rispetto alla variabile  $X$  è la retta  $y = mx + q$  per la quale risulta minima la somma dei quadrati degli scarti tra i valori rilevati  $y_i$  e i corrispondenti valori sulla retta  $f(x_i) = mx_i + q$ . Si può dimostrare che i coefficienti  $m$  e  $q$  di tale retta sono dati dalle seguenti formule

$$m = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad \left( o, \text{ equivalentemente, } m = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2} \right)$$

$$q = \bar{y} - m\bar{x}$$

ove  $\overline{xy}$  rappresenta la media dei prodotti  $x_i \cdot y_i$  e  $\overline{x^2}$  la media dei quadrati  $x_i^2$ .

Il coefficiente di correlazione campionaria (o coefficiente di correlazione lineare di Pearson) è definito come

$$r := \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}} \quad \left( o, \text{ equivalentemente, } r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{(\overline{x^2} - \bar{x}^2)(\overline{y^2} - \bar{y}^2)}} \right).$$

Tale coefficiente è un numero compreso fra -1 e 1 e misura la bontà dell'interpolazione fornita dalla retta di regressione: se  $r$  è sufficientemente vicino a 1 o a -1 (almeno 0.8 in valore assoluto), allora l'interpolazione fornita dalla retta di regressione è buona; se invece  $r$  è prossimo a 0, i dati non seguono affatto un andamento lineare.

Es. 1) In un gruppo di 5 adulti, la somministrazione di dosi di un farmaco ha comportato le seguenti diminuzioni della pressione arteriosa:

<i>Dose (in mg)</i>	<i>Diminuzione di pressione (in mmHg)</i>
7	10
12	18
15	20
20	25
22	25

- a) Riportare i dati in un diagramma a dispersione;  
b) Scrivere l'equazione della retta di regressione e disegnarla;  
c) Calcolare il coefficiente di correlazione campionaria e dedurre se i dati hanno un andamento lineare;  
d) Calcolare, se possibile, la dose ottimale per ottenere una diminuzione della pressione di 15 mmHg.

(Soluz:  $y = x + 4$ , dose ottimale  $\simeq 11$  mg.)

Es. 2) In uno studio sulla relazione tra l'età e il peso delle cavie di laboratorio sono stati rilevati i seguenti dati

<i>Età(X)(in giorni)</i>	61	76	80	66	71	68	78	55	74	60	65	70	64	73	68
<i>Peso (Y) (in grammi)</i>	28	32	37	29	31	30	32	26	32	27	29	30	28	31	31

Calcolare la retta di regressione e il coefficiente di correlazione lineare. Interpretare i risultati ottenuti.  
Calcolare, se possibile, il peso previsto per una cavia di 50 giorni.

(Soluz:  $y = 0.35x + 6.14$ ;  $r \simeq 0.927$ , dunque c'è buona correlazione lineare tra i dati;...)