

Interdipendenza

- entriamo nel campo dei **caratteri quantitativi...**
- variabili numeriche (come reddito, altezza, età, ecc.)
- studiamo **l'interdipendenza**, cioè **quanto e come due variabili variano insieme**

Interdipendenza

- Quando studiamo due **caratteri quantitativi** X e Y , vogliamo capire **se e come variano insieme**
- Esempio: Ore di studio vs Voto esame

Interdipendenza

Per ogni coppia di unità statistiche (i, j) con valori (x_i, y_i) e (x_j, y_j)

- Le due coppie sono **concordanti** se:

$$(x_i - x_j)(y_i - y_j) > 0$$

Significa che, quando X aumenta, anche Y aumenta (o entrambi diminuiscono).

Interdipendenza

Per ogni coppia di unità statistiche (i, j) con valori (x_i, y_i) e (x_j, y_j)

- Le due coppie sono **discordanti** se:

$$(x_i - x_j)(y_i - y_j) < 0$$

Significa che, quando X aumenta, Y diminuisce

Interdipendenza

Per ogni coppia di unità statistiche (i, j) con valori (x_i, y_i) e (x_j, y_j)

- Le due coppie non sono **ne discordanti ne concordanti** se:

$$(x_i - x_j)(y_i - y_j) = 0$$

Interdipendenza

- La prevalenza di coppie **concordanti** indica una **correlazione positive**
- la prevalenza di **discordanti** indica una **correlazione negativa**

Covarianza

- La **covarianza** misura come due variabili quantitative variano insieme
- Sia una popolazione di n unità con valori (x_i, y_i) .
Si calcolano le medie:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad e \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

Covarianza

- La **covarianza** è definita come:

$$\text{COV}(X, Y) = \frac{1}{n} \sum_{i=1}^n (x_i - x_m)(y_i - y_m)$$

Interpretazione

- **associazione diretta (positiva)** $\Rightarrow$ al crescere di X , tende a crescere anche Y
- **associazione inversa (negativa)** $\Rightarrow$ al crescere di X , Y tende a diminuire
- **assenza di relazione lineare** (ma può esserci una relazione non lineare)

Covarianza

- La covarianza dipende dalle **unità di misura** (es. cm, kg, euro), quindi **non è standardizzata** e non è confrontabile tra variabili diverse.
- Per ovviare a questo, si usa un indice **adimensionale**:

il **coefficiente di correlazione lineare di Bravais e Pearson**

Covarianza

$$\rho_{xy} = \frac{Cov(X, Y)}{S_x * S_y}$$

$$s_x = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

$$s_y = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}$$

Range dei valori

$$-1 \leq r_{xy} \leq +1$$

$r = +1 \rightarrow$ correlazione lineare **positiva perfetta** (i punti stanno su una retta crescente)

$r = -1 \rightarrow$ correlazione lineare **negativa perfetta** (i punti stanno su una retta decrescente)

$r = 0 \rightarrow$ **assenza di correlazione lineare**

Covarianza

Supponiamo di osservare 5 individui:

i	X_i (ore di studio)	Y_i (voto)
1	2	60
2	3	65
3	4	70
4	5	75
5	6	80

Recap

BES: Titolo di studio vs partecipazione culturale (dominio Istruzione e Cultura).

Regione	Laureati (%)	Partecipazione culturale (%)
A	18	28
B	21	35
C	24	41
D	27	43

Esercizi

1. Trasforma le percentuali in valori (immagina 10.000 abitanti per regione).
2. Costruisci le **frequenze congiunte** per classi di titolo di studio (Basso/Medio/Alto) e partecipazione culturale (Bassa/Media/Alta).
3. Calcola le **frequenze marginali** e discuti quali regioni mostrano maggiore accesso alla cultura.

Recap

Regione	Laureati (%)	Partecipazio ne culturale (%)
A	18	28
B	21	35
C	24	41
D	27	43

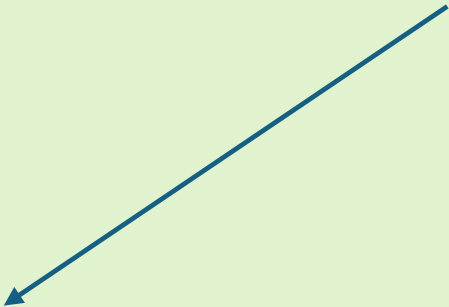
Classi:

Laureati: Basso (≤ 20), Medio (21–24), Alto (≥ 25)

Partecipazione: Basso (≤ 30), Media (31–38), Alta (≥ 39)

Procediamo con l'assegnazione per regione...

	Bassa	Media	Alta	Totale
Basso	10000	0	0	10000
Medio	0	10000	10000	20000
Alto	0	0	10000	10000
Totale	10000	10000	20000	40000



Recap

BES: Soddisfazione per la vita vs condizione occupazionale.

Condizione	Soddisfatto	Non soddisfatto	Totale
Occupato	350	150	500
Disoccupato	120	180	300
Inattivo	100	100	200
Totale	570	430	1000

Esercizi

1. Calcola le **frequenze attese**.
2. Calcola la **statistica χ^2** .
3. Decidi se soddisfazione e condizione occupazionale sono **indipendenti**.
4. Calcola il **coefficiente V di Cramer** e interpreta la forza dell'associazione.

Recap

	Condizione	Soddisfatto	Non soddisfatto	Totale
	Occupato	350	150	500
	Disoccupato	120	180	300
	Inattivo	100	100	200
	Totale	570	430	1000

Attese $E_{ij} = \frac{n_{i.}n_{.j}}{n}$:

Occ-Sod: 285; Occ-Non: 215

Dis-Sod: 171; Dis-Non: 129

Ina-Sod: 114; Ina-Non: 86

Chi-quadro $\chi^2 = \sum \frac{(O-E)^2}{E} \approx 73.847$

Cramer V = 0.272 → **associazione moderata**

Recap

BES: Distribuzione del reddito equivalente per classi.

Classe di reddito (€)	Frequenza n_i
0 – 10.000	40
10.001 – 20.000	30
20.001 – 30.000	20
> 30.000	10
Totale	100

Esercizi

1. Calcola il **reddito medio di classe**.
2. Calcola le **quote di reddito** e le **cumulate**.
3. Usa la formula del Gini per dati raggruppati
4. Commenta il grado di disuguaglianza.

Recap

Classe di reddito (€)	Frequenza n_i
0 – 10.000	40
10.001 – 20.000	30
20.001 – 30.000	20
> 30.000	10
Totale	100

- Facciamo finta che tutti gli individui in una classe guadagnino esattamente il valore medio di quella classe (o che il totale del reddito della classe sia distribuito equamente)
- Se non abbiamo il totale, lo stimiamo moltiplicando il **valore centrale della classe** per la frequenza

Recap

Classe di reddito (€)	Frequenza n_i
0 – 10.000	40
10.001 – 20.000	30
20.001 – 30.000	20
> 30.000	10
Totale	100

Per ogni classe:

$$m_i = \frac{\text{limite inferiore} + \text{limite superiore}}{2}$$

esclusa l'ultima classe aperta, dove si sceglie un valore plausibile

Recap

Classe (€)	n_i	m_i	$n_i \cdot m_i$
0-10.000	40	5.000	200.000
10.001-20.000	30	15.000	450.000
20.001-30.000	20	25.000	500.000
>30.000	10	35.000	350.000
Totale	100	—	1.500.000

Per ogni classe:

$$m_i = \frac{\text{limite inferiore} + \text{limite superiore}}{2}$$

esclusa l'ultima classe aperta, dove si sceglie un valore plausibile

Recap

Classe (€)	n_i	m_i	$n_i \cdot m_i$
0-10.000	40	5.000	200.000
10.001-20.000	30	15.000	450.000
20.001-30.000	20	25.000	500.000
>30.000	10	35.000	350.000
Totale	100	—	1.500.000

Frequenze relative:

$$f_i = \frac{n_i}{N}$$
$$f_1 = 0.40, f_2 = 0.30, f_3 = 0.20, f_4 = 0.10$$

Quote di reddito:

$$p_i = \frac{n_i m_i}{\sum n_i m_i}$$
$$p_1 = \frac{200000}{1500000} = 0.1333$$
$$p_2 = \frac{450000}{1500000} = 0.3000$$
$$p_3 = \frac{500000}{1500000} = 0.3333$$
$$p_4 = \frac{350000}{1500000} = 0.2333$$

Recap

Calcola le cumulate per le frequenze F_i e per le quote di reddito P_i

Classe	f_i	F_i	p_i	P_i
1	0.40	0.40	0.1333	0.1333
2	0.30	0.70	0.3000	0.4333
3	0.20	0.90	0.3333	0.7666
4	0.10	1.00	0.2333	1.0000

Recap

$$G=1-i=1\sum_k (P_i+P_{i-1}) (F_i-F_{i-1})$$

i	<i>P_{i-1}</i>	<i>P_i</i>	<i>F_{i-1}</i>	<i>F_i</i>	<i>(P_i + P_{i-1})(F_i - F_{i-1})</i>
1	0.0000	0.1333	0.0000	0.40	0.0533
2	0.1333	0.4333	0.40	0.70	0.1700
3	0.4333	0.7666	0.70	0.90	0.2400
4	0.7666	1.0000	0.90	1.00	0.1767

Recap

Classe di reddito (€)	Frequenze
0–10.000	25%
10.001–20.000	35%
20.001–30.000	25%
>30.000	15%

Medie di classe (scelte): 5k, 15k, 25k, 35k.

Quote di reddito $p_i = \frac{n_i m_i}{\sum n_k m_k}$ → cumulate P_i e cumulate frequenze F_i

Risultato: $G \approx 0.308$ **disuguaglianza moderata**

Recap

BES: Benessere economico

Regione	PII pro capite (€)	Laureati (%)
A	22.000	18
B	25.000	21
C	29.000	24
D	33.000	28

Esercizi

1. Calcola le medie $\bar{x}, \bar{y}$.
2. Calcola **Cov(X,Y)**.
3. Calcola **Pearson r**.
4. Interpreta la relazione. È forte? Lineare? Positiva?

Recap

BES: Benessere economico

Regione	PIL pro capite (€)	Laureati (%)
A	22.000	18
B	25.000	21
C	29.000	24
D	33.000	28

Risultati: $Cov \approx 15312.5$, $r \approx 0.998 \rightarrow$ correlazione lineare fortissima positiva

Recap

BES: Valutazione benessere psicologico (1-5) e soddisfazione lavorativa (1-5) per 6 individui:

Persona	Psicologico	Lavoro
1	1	2
2	2	3
3	3	3
4	3	4
5	4	4
6	5	5

Esercizi:

Quanto è forte la relazione? Calcola Rho di Spearman

Recap

Siccome i valori sono ordinali: i ranghi = valori, con un caso di pareggio

Persona	X	rango X	Y	rango Y	d = Rx - Ry
1	1	1	2	1.5	-0.5
2	2	2	3	2.5	-0.5
3	3	3.5	3	2.5	1.0
4	3	3.5	4	4.5	-1.0
5	4	5	4	4.5	0.5
6	5	6	5	6	0