

Rapine e Disoccupazione: un approccio statistico

Mattia Gallucci



Contenuti

01.

Raccolta dati

Frequenze

02.

Indici di posizione

Media, mediana, moda

03.

Indici di variabilità

Varianza, Deviazione
Standard, SMA, Ampiezza e
Coefficiente di Variazione.

04.

Indici di forma

Indice di Asimmetria e di
Curtosi.

05.

Percentili

Percentili campionari, Box
Plot e dati bivariati.

01.

Raccolta dati

Frequenze



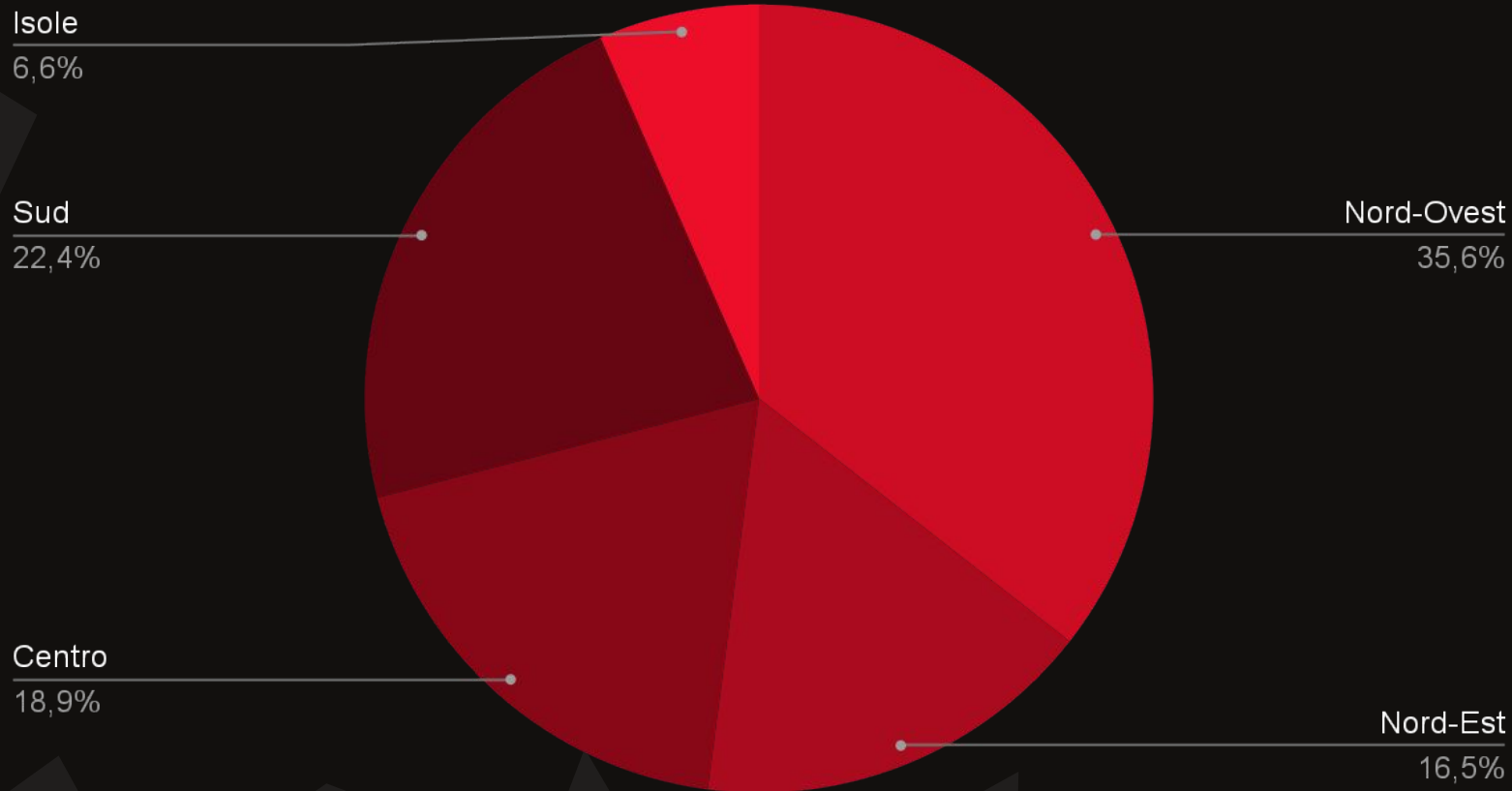
Dati

| Regione | # Rapine | Disoccupati (migliaia) |
|---------------------|----------|------------------------|
| Piemonte | 2 060 | 124 |
| Valle D'Aosta | 12 | 3 |
| Liguria | 842 | 46 |
| Lombardia | 6205 | 225 |
| Trentino Alto Adige | 367 | 16 |
| Veneto | 1570 | 94 |
| Friuli | 231 | 29 |
| Emilia Romagna | 2067 | 105 |
| Toscana | 1767 | 104 |
| Umbria | 168 | 27 |

| Regione | # Rapine | Disoccupati (migliaia) |
|------------|----------|------------------------|
| Marche | 267 | 42 |
| Lazio | 2640 | 194 |
| Abruzzo | 242 | 50 |
| Molise | 30 | 12 |
| Campania | 4137 | 339 |
| Puglia | 1081 | 174 |
| Basilicata | 54 | 14 |
| Calabria | 211 | 80 |
| Sicilia | 1439 | 265 |
| Sardegna | 252 | 74 |
| Totale | 25642 | 2027 |

Fonte: ISTAT 2022

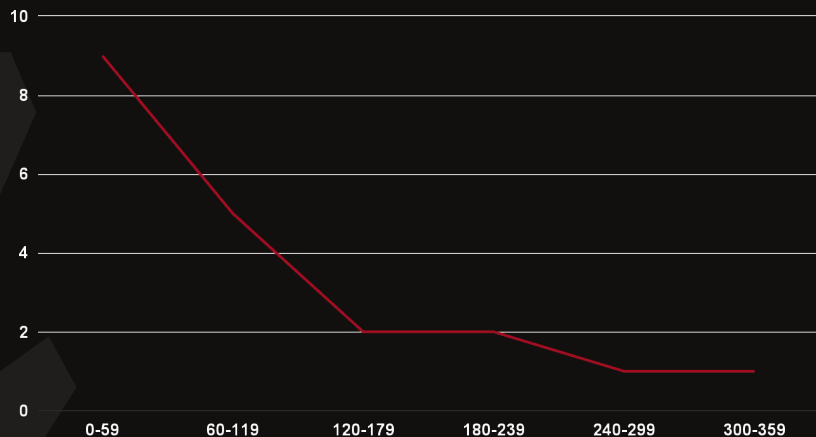
Percentuale Rapine per Aree Geografiche



Frequenza Assoluta

f_i : frequenza assoluta della modalità vi , ossia il **numero di dati** del campione (x_1, \dots, x_n) che hanno valore vi , per $i = 1, 2, \dots, k$

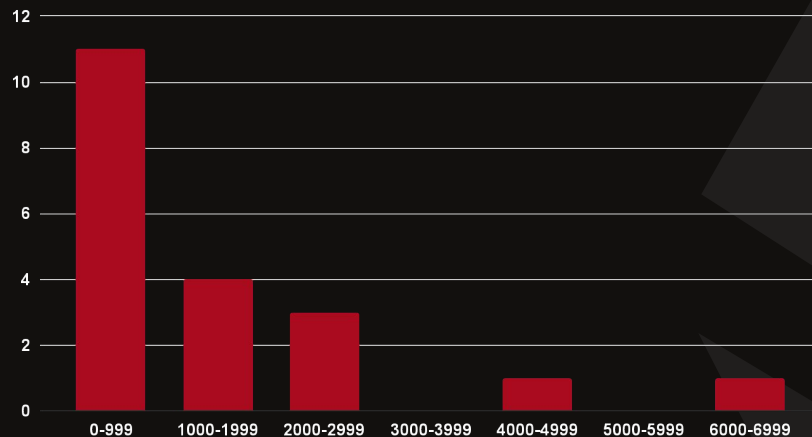
Frequenza Assoluta



DISOCCUPATI

Frequenza Assoluta

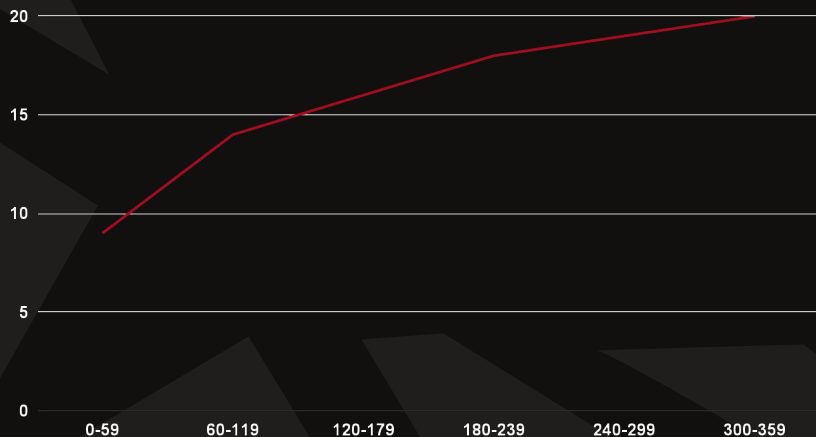
RAPINE



Frequenza Cumulativa Assoluta

F_i : frequenza cumulativa assoluta delle modalità v_1, \dots, v_i , ossia **numero di dati** del campione (x_1, \dots, x_n) che hanno **valore minore o uguale a v_i** , per $i = 1, 2, \dots, k$

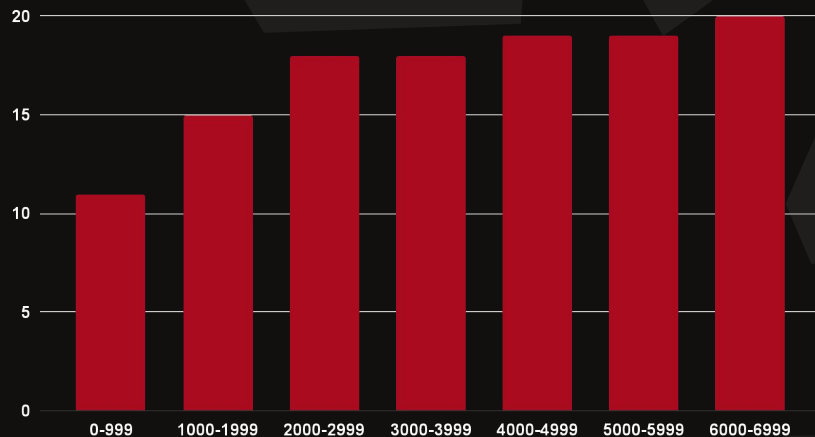
Frequenza Cumulativa Assoluta



DISOCCUPATI

Frequenza Cumulativa Assoluta

RAPINE

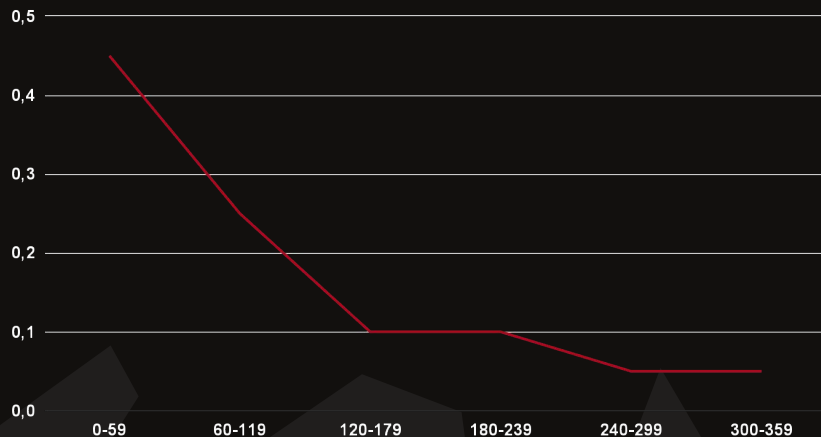


$$F_i = \sum_{j=1}^i f_j$$

Frequenza Relativa

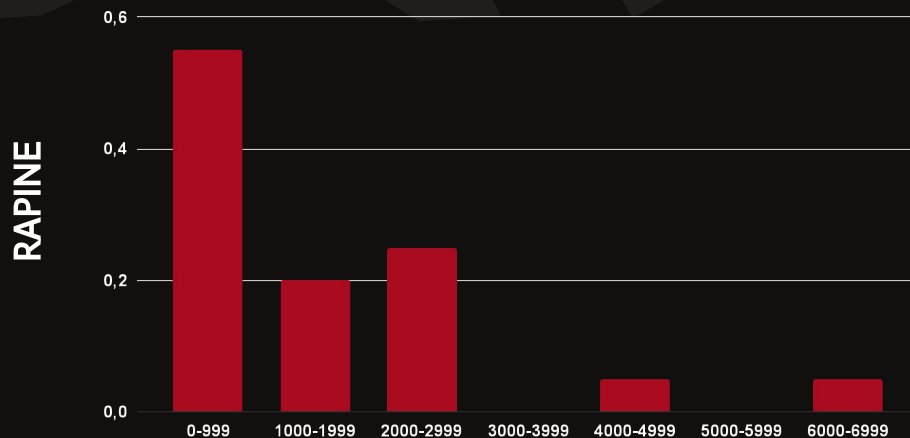
p_i : frequenza relativa della modalità v_i , ossia **rapporto** tra la frequenza assoluta f_i e l'ampiezza n del campione, per $i = 1, 2, \dots, k$

Frequenza Relativa



DISOCCUPATI

Frequenza Relativa

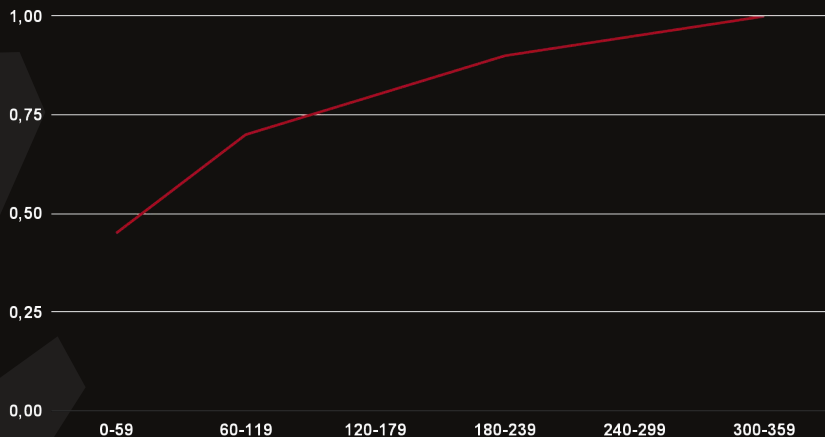


$$p_i = \frac{f_i}{n}$$

Frequenza Cumulativa Relativa

P_i : frequenza cumulativa relativa delle modalità v_1, \dots, v_i , ossia **somma delle frequenze relative** delle modalità v_1, \dots, v_i , per $i = 1, 2, \dots, k$

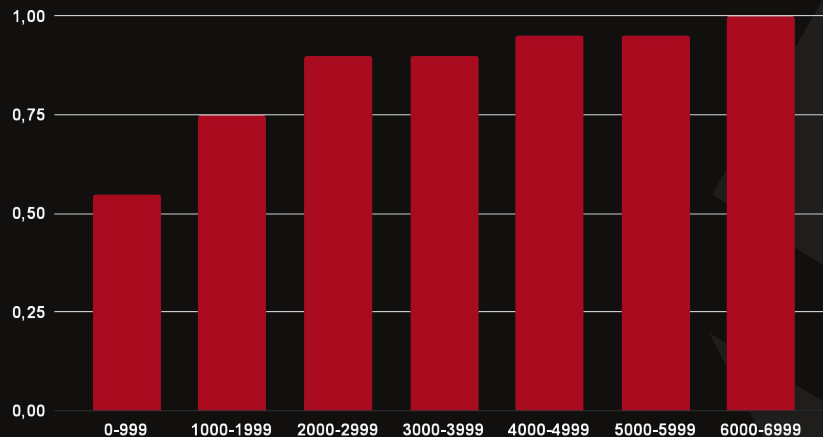
Frequenza Cumulativa Relativa



DISOCCUPATI

Frequenza Cumulativa Relativa

RAPINE



$$P_i = \sum_{j=1}^i p_j = \frac{F_i}{n}$$



02.

Indici di posizione

Media, mediana, moda

Indici di Posizione

Media

Casi di rapine in media in una regione: 1282.1, quindi circa **1282** rapine. Molto maggiore rispetto alla media di 100.85 disoccupati, quindi circa **101**

Mediana

Valore centrale del campione dei dati: **604.5** per le rapine e **77** per la disoccupazione

Moda

La classe con frequenza massima è: **0-999** per le classi riguardanti le rapine, **0-99** per le classi di disoccupazione



03.

Indici di Variabilità

Varianza, Deviazione Standard, SMA, Ampiezza e Coefficiente di Variazione.



Indici di Variabilità

| | VARIANZA | DEVIAZIONE STANDARD | SCARTO MEDIO ASSOLUTO | AMPIEZZA DEL CAMPIONE | COEFFICIENTE DI VARIAZIONE |
|----------------|--|---------------------|--|-------------------------|----------------------------|
| Rapine | 2411021.09 | 1552.7 | 1140.8 | 6193 | 1.2 |
| Disoccupazione | 8337.6275 | 91.3 | 72.32 | 336 | 0,9 |
| | $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ | $s = \sqrt{s^2}$ | $s_a = \frac{1}{n} \sum_{i=1}^n x_i - \bar{x} $ | $w = x_{(n)} - x_{(1)}$ | $cv = \frac{s}{\bar{x}}$ |

04. **Indici di forma**

Indice di Asimmetria e di Curtosi.



Indici di Forma

Gli indici di forma misurano **caratteristiche** relative alla forma della **distribuzione dei dati**

Indice di Asimmetria

$$\gamma_R = \frac{1}{s^3} \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3 = 1,76865$$

$$\gamma_D = 1,08256$$

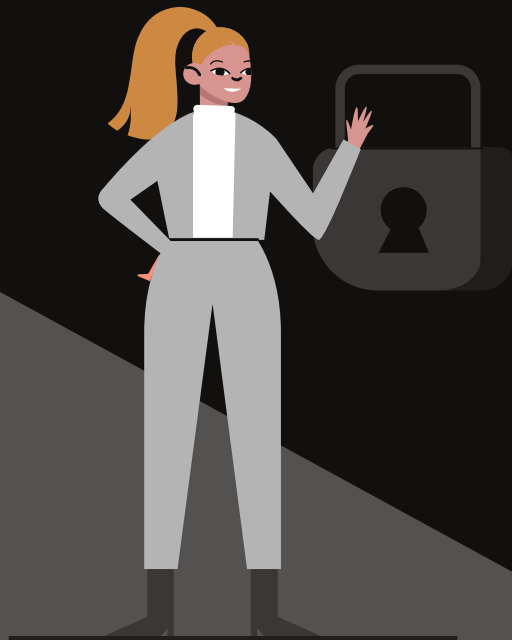
Siccome γ è positivo, le distribuzione di dati presentano una **coda più lunga a destra**

Indice di Curtosi

$$k_R = \left(\frac{1}{s^4} \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4 \right) - 3 = 2,79356$$

$$k_D = 0,352347$$

Essendo k positivo, è presente un eccesso di dati nelle classi centrali. I dati hanno distribuzione **leptocurtica**



05.

Percentili

Percentili campionari, Box Plot
e dati bivariati.

Percentili

Un percentile k-esimo di un campione di dati è un **valore** che è **maggiore** di una percentuale **k** dei dati e **minore della restante percentuale**

25% - 1° Quartile

Rapine: 216
Disoccupati: 27,5

50% - 2° Quartile

Rapine: 604,5
Disoccupati: 77

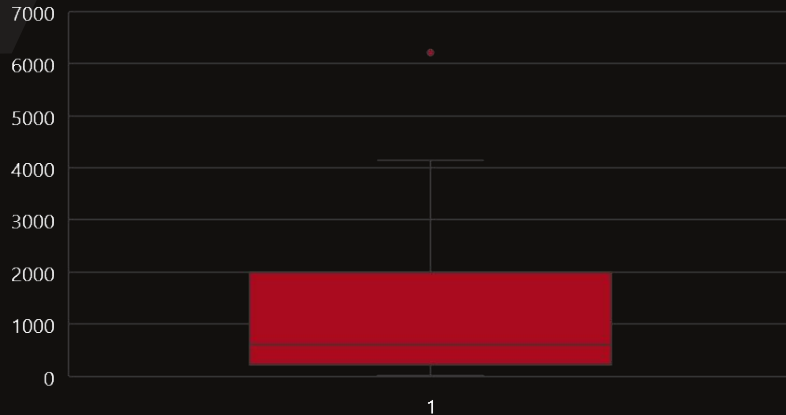
75% - 3° Quartile

Rapine: 1986,75
Disoccupati: 161,5

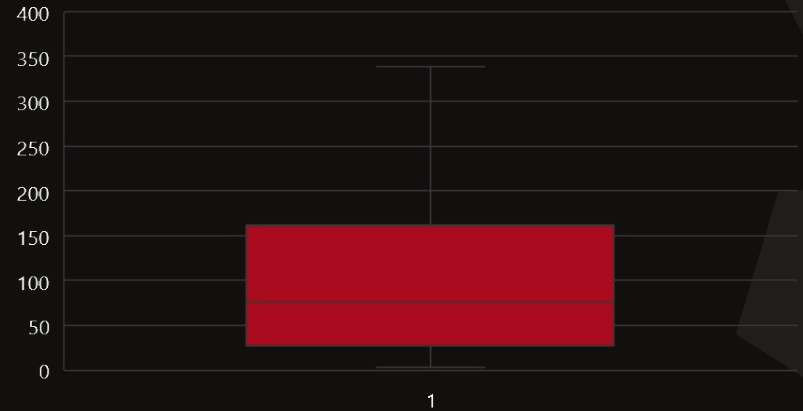
Box Plot

Per visualizzare i quartili e la loro differenza, è utile impiegare un Box Plot. Questo grafico mette in risalto le **distanze tra i quartili** in modo chiaro ed efficace.

Box Plot Rapine



Box Plot Disoccupati



Disuguaglianza di Chebyshev

Fornisce un utile **limite inferiore** per il numero di dati vicini alla media campionaria

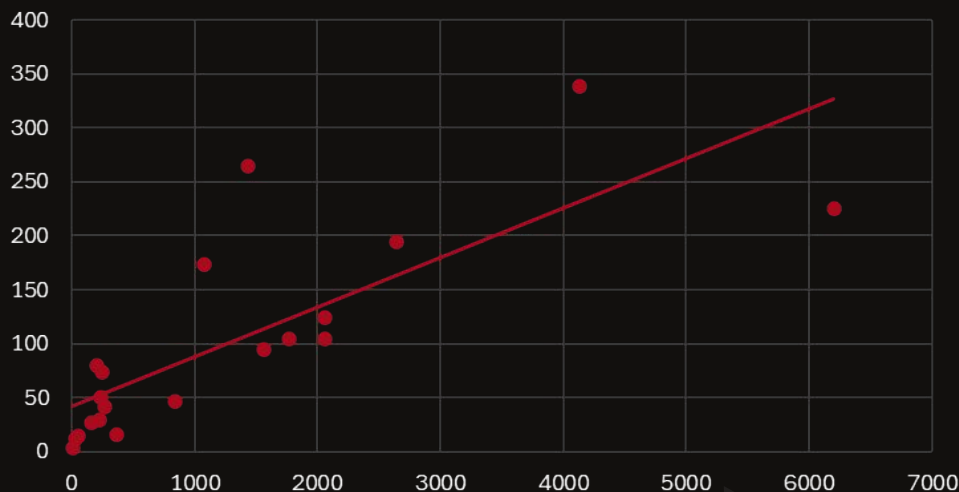
$$\frac{|S_k|}{n} > \begin{cases} 0,55 & k = 1, 5 \\ 0,75 & k = 2 \\ 0,88 & k = 3 \end{cases}$$

- All'intervallo $(\bar{x} - 1,5s, \bar{x} + 1,5s)$ è almeno pari al 55,56%;
 - Nel campione delle rapine, tra (-1047, 3611), è **90%**
- All'intervallo $(\bar{x} - 2s, \bar{x} + 2s)$ è almeno pari al 75%;
 - Nel campione delle rapine, tra (-1823, 4388), è **95%**
- All'intervallo $(\bar{x} - 3s, \bar{x} + 3s)$ è almeno pari all'88,89%;
 - Nel campione delle rapine, tra (-3376, 5940), è **95%**

Dati Bivariati

Per esplorare questo aspetto più approfonditamente, se esiste una **possibile relazione** tra il numero di rapine e il tasso di disoccupazione, potremmo utilizzare i dati di disoccupazione delle regioni e visualizzarli tramite un **diagramma a dispersione**

Diagramma a Dispersione



Equazione retta di regressione: $y = 42 + 0,05x$

Guardando al diagramma, è evidente come l'**aumento della disoccupazione** sia accompagnato da un **aumento nel numero di rapine**. Questo non solo suggerisce una correlazione logica ovvia, ma anche un coefficiente di correlazione che tende verso l'unità.

Coefficiente di Correlazione

Infine, per concludere la nostra indagine statistica, procederemo al calcolo del coefficiente di correlazione. Questo ci permetterà di **valutare la relazione** tra due set di dati, nel nostro caso, il numero di rapine e il numero di disoccupati.



$$r = \frac{1}{(n - 1)s_x s_y} \sum_{i=1}^n (x_i - \bar{x}) (y_i - \bar{y}) = 0,78$$

In conclusione, si può osservare (come già evidenziato dal grafico) che il numero di rapine e il numero di disoccupati mostrano una **correlazione diretta e proporzionale**, come indicato da un coefficiente di correlazione vicino all'unità.

Questo studio ha evidenziato una **forte correlazione** tra il numero di rapine e il tasso di disoccupazione, sottolineando l'influenza dei **fattori socio-economici** sulla criminalità locale. L'aumento della disoccupazione è associato a un incremento delle rapine, indicando un legame diretto tra difficoltà economiche e livelli di criminalità. Questi risultati suggeriscono che interventi mirati per **migliorare l'occupazione** e sostenere l'economia potrebbero **ridurre** efficacemente la **criminalità** urbana, fornendo un approccio strategico per affrontare le sfide di sicurezza e sviluppo nelle comunità.



Grazie per l'attenzione!

Mattia Gallucci

