

La Criminalità in Italia: Relazioni con Istruzione, Disoccupazione e Reddito

Andrea Marcuzzi

2025-03-25

Presentata da: Andrea Marcuzzi
Andrea Marcuzzi
Matricola:

Relatore:
Prof Paolo Verme

ANNO ACCADEMICO 24 / 25

1 Introduzione

La criminalità rappresenta un tema centrale nell'analisi delle dinamiche sociali ed economiche di un Paese. Comprendere quali fattori possano influenzarla è cruciale per sviluppare politiche efficaci di prevenzione e sicurezza.

In questo elaborato si intende analizzare il fenomeno della criminalità in Italia nel periodo 2010–2023 e la sua possibile relazione con alcune variabili socioeconomiche rilevanti, quali:

- il livello di istruzione della popolazione,
- il tasso di disoccupazione,
- il reddito disponibile delle famiglie.

L'obiettivo è valutare se e in che misura questi fattori siano correlati all'andamento della criminalità nel tempo, utilizzando strumenti di analisi statistica su dati ufficiali rilasciati da ISTAT.

L'analisi si basa su una pipeline di lavoro riproducibile in ambiente RStudio, con l'utilizzo combinato di `Quarto`, `ggplot2`, `dplyr` e pacchetti API (`rsdmx`) per l'estrazione automatica dei dati.

2 Obiettivi

L'analisi si propone di:

1. Descrivere l'evoluzione della criminalità in Italia dal 2010 al 2023.
2. Analizzare i trend socioeconomici relativi a istruzione, disoccupazione e reddito.
3. Valutare l'esistenza di correlazioni tra criminalità e ciascuna variabile indipendente.
4. Stimare semplici modelli di regressione lineare per individuare relazioni statisticamente significative.
5. Riflettere sul significato dei risultati ottenuti in un'ottica interpretativa e di policy.

3 Reperimento Dati

I dati utilizzati in questa analisi sono stati reperiti principalmente tramite il portale ufficiale dell'ISTAT (Istituto Nazionale di Statistica), facendo uso del protocollo SDMX (Statistical Data and Metadata eXchange).

L'accesso ai dati è avvenuto tramite API REST messe a disposizione dall'ISTAT, grazie all'utilizzo del pacchetto `rsdmx` in R. Questo ha consentito di interrogare i dataflow ufficiali e scaricare automaticamente dataset aggiornati in formato strutturato e leggibile.

Le categorie di dati utilizzate sono:

- Criminalità: numero di detenuti classificati per tipo di reato, sesso e cittadinanza (`flowRef = 73_173_DF_DCCV_DETENUTI_1`);
- Istruzione: popolazione per livello di istruzione più elevato raggiunto, filtrata per età 15-64 (`flowRef = 150_1190_DF_DCCV_FORZLV1_UNT2020_1`);
- Disoccupazione: tasso di disoccupazione medio annuo (`flowRef = 151_914`);
- Reddito disponibile: valori annuali del reddito disponibile delle famiglie italiane (`flowRef = 737_1093`).

Tutti i dati sono di tipo open data, accessibili liberamente e rilasciati da fonte ufficiale. È stata eseguita una fase di pulizia e aggregazione per armonizzare i dataset per anno (`obsTime`) e variabile (`obsValue`) al fine di permettere analisi comparative coerenti.

I dati sono stati successivamente salvati in formato `.csv` e integrati nel workflow tramite script R automatizzati.

4 Preprocessing dei Dati

Dopo il reperimento, i dati sono stati sottoposti a un'attenta fase di preprocessing, volta a renderli compatibili per l'analisi esplorativa e statistica. Le operazioni effettuate includono:

- Selezione delle variabili rilevanti (`obsTime`, `obsValue`) da ciascun dataset;
- Filtraggio temporale per il periodo 2010–2020, ove non possibile sono stati presi solamente anni comuni dei dataset da confrontare;
- Aggregazione annuale dei dati trimestrali, ove necessario (es. disoccupazione);
- Sommatoria dei valori nel caso di dati disaggregati per età, genere o territorio (es. crimini per tipo e cittadinanza);
- Gestione dei valori mancanti mediante imputazione con la media nei casi limitati (es. reddito);
- Uniformazione della codifica temporale (es. conversione da “2010-Q1” a “2010”).

Tutti i dataset sono stati armonizzati sul campo `obsTime`, per permettere il merge e l'analisi congiunta.

Esempio: per il dataset della disoccupazione, sono stati selezionati solo i dati relativi alla fascia di età 15–64 anni e successivamente è stata calcolata la media annuale dei tassi trimestrali.

Infine, ciascun dataset è stato esportato in formato `.csv` come base coerente per la successiva fase di analisi.

5 Analisi Dati e Grafici

5.1 Analisi Descrittiva della Criminalità

L'analisi descrittiva dei dati sulla criminalità consente di comprendere le principali dinamiche del fenomeno in Italia durante il periodo considerato. Innanzitutto, è stato calcolato il numero totale di reati per anno, ottenuto mediante l'aggregazione dei dati originali per anno di osservazione (`obsTime`). Questo aggregato rappresenta un indicatore sintetico dell'andamento della criminalità nel tempo.

Successivamente, i reati sono stati analizzati per tipologia (`TYPE_OFFENCE`), al fine di identificare le categorie di reato più diffuse. I risultati mostrano che i reati più comuni rientrano nelle categorie generali (es. `ALL`), ma includono anche tipologie specifiche come furti (`FU`), rapine (`RA`), violenza (`VIR`) e lesioni personali (`OMV`), suggerendo che la componente predatoria e quella violenta costituiscono una parte consistente del fenomeno criminale.

È stata inoltre effettuata un'analisi per genere degli autori dei reati (`SEX`), che evidenzia una netta prevalenza maschile, coerente con i dati storici e letteratura criminologica. La componente femminile e quella “non classificata” risultano molto più contenute.

Infine, i dati sono stati suddivisi per cittadinanza (`CITIZENSHIP`), distinguendo tra cittadini italiani, stranieri e valori aggregati. Anche in questo caso, la maggior parte dei reati è attribuita a cittadini italiani, ma si registra una quota non trascurabile associata a soggetti di cittadinanza straniera, dato che potenzialmente apre a riflessioni di natura socio-economica e territoriale.

Complessivamente, questi riepiloghi forniscono una base solida per comprendere il contesto generale della criminalità, e per individuare possibili fattori associati da approfondire nelle sezioni successive.

```
# A tibble: 14 x 2
  obsTime Totale_Criminalità
  <chr>      <dbl>
1 2010      863634.
2 2011      875875.
3 2012      880665.
4 2013      861199.
5 2014      766164.
```

6	2015	747802.
7	2016	773227.
8	2017	811362.
9	2018	841235.
10	2019	870425.
11	2020	790003.
12	2021	802853.
13	2022	837459.
14	2023	894409.

A tibble: 54 x 2

	TYPE_OFFENCE	Numero_Casi
	<chr>	<dbl>
1	ALL	3301708
2	STU	1208912.
3	RA	968640.
4	FU	723656.
5	RIC	634280.
6	LPV	596375.
7	OMV	502916.
8	ES	434129.
9	VIR	430226.
10	AFS	406292.

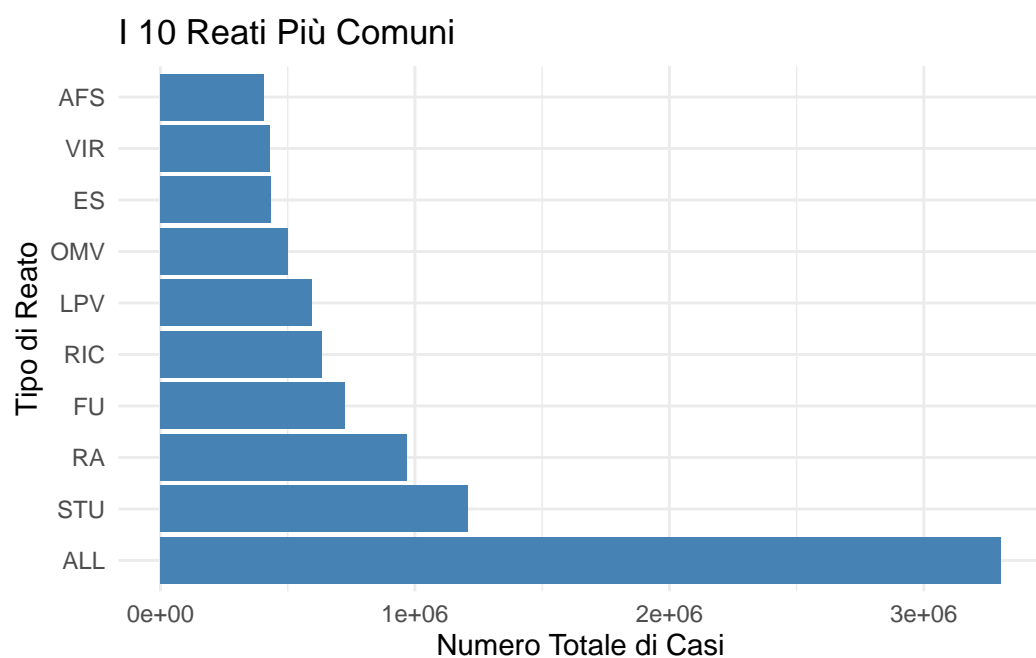
i 44 more rows

A tibble: 3 x 2

	SEX	Totale_Criminalità
	<chr>	<dbl>
1	1	5594920.
2	2	217410.
3	9	5803981.

A tibble: 3 x 2

	CITIZENSHIP	Totale_Criminalità
	<chr>	<dbl>
1	IT	4128313.
2	WORLD	5803624.
3	WRL_X_ITA	1684374.



5.2 Analisi del Livello di Istruzione

L'istruzione rappresenta uno dei principali indicatori del capitale umano di una popolazione, nonché un possibile fattore determinante della criminalità. In questa sezione si

analizza il totale della popolazione con età compresa tra i 15 e i 64 anni, suddivisa per livello di istruzione più elevato raggiunto.

Il trend complessivo mostra una crescita graduale e costante nel numero di persone con un titolo di studio registrato. Questo può essere attribuito all'espansione dell'accesso all'istruzione superiore, alla maggiore permanenza nei percorsi scolastici e alla diffusione di politiche di contrasto alla dispersione scolastica.

È importante sottolineare che, sebbene il dato assoluto sia in aumento, l'effetto dell'istruzione sulla criminalità richiede un'analisi più fine. Infatti, la variabilità annuale è contenuta e l'effetto potenziale potrebbe dipendere dalla distribuzione per tipologia di titolo (basso, medio, alto), più che dal volume aggregato.

L'incremento del livello medio di istruzione può teoricamente contribuire alla riduzione del tasso di criminalità, in quanto l'istruzione è spesso associata a migliori opportunità occupazionali, maggiore integrazione sociale e minore esposizione a comportamenti devianti. Tuttavia, come vedremo più avanti, la relazione osservata risulta moderatamente negativa, ma non statisticamente significativa.

```
# A tibble: 5 x 2
  EDU_LEV_HIGHEST Totale_Persone
      <int>         <dbl>
1          99      87697872.
2           7      24787493.
3           4      20819724.
4          11      8724363.
5           3       7875557.

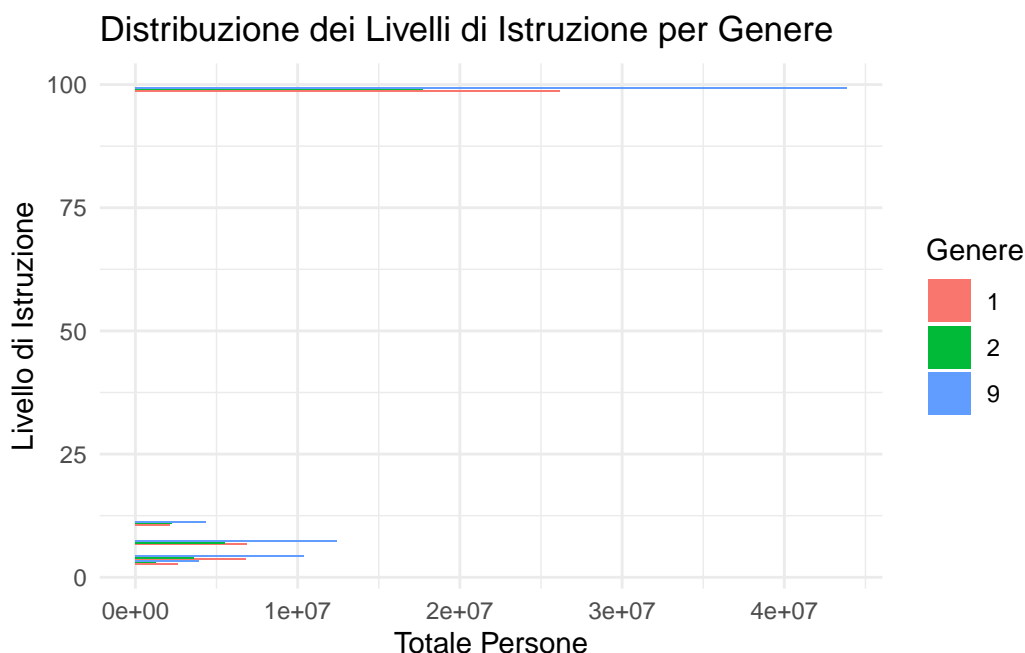
# A tibble: 1,100 x 3
# Groups:   obsTime [220]
  obsTime EDU_LEV_HIGHEST Totale_Persone
    <chr>         <int>         <dbl>
1 1977             3         57594.
2 1977             4         28067.
3 1977             7         14086.
4 1977            11          4316.
5 1977            99        224588.
6 1977-Q1          3         58267.
7 1977-Q1          4         27181.
8 1977-Q1          7         13786.
9 1977-Q1         11          4275.
10 1977-Q1         99        223331.
# i 1,090 more rows
```

A tibble: 15 x 3

Groups: SEX [3]

	SEX	EDU_LEV_HIGHEST	Totale_Persone
	<int>	<int>	<dbl>
1	9	99	43848936.
2	1	99	26156326.
3	2	99	17692610.
4	9	7	12393746.
5	9	4	10409862.
6	1	7	6887512.
7	1	4	6815714.
8	2	7	5506234.
9	9	11	4362181.
10	9	3	3937778.
11	2	4	3594148.
12	1	3	2643583.
13	2	11	2229852.
14	1	11	2132329.
15	2	3	1294196.

[1]	"FREQ"	"FREQ_label.en"
[3]	"FREQ_label.it"	"REF_AREA"
[5]	"REF_AREA_label.en"	"REF_AREA_label.it"
[7]	"DATA_TYPE"	"DATA_TYPE_label.en"
[9]	"DATA_TYPE_label.it"	"SEX"
[11]	"SEX_label.en"	"SEX_label.it"
[13]	"AGE"	"AGE_label.en"
[15]	"AGE_label.it"	"EDU_LEV_HIGHEST"
[17]	"EDU_LEV_HIGHEST_label.en"	"EDU_LEV_HIGHEST_label.it"
[19]	"CITIZENSHIP"	"CITIZENSHIP_label.en"
[21]	"CITIZENSHIP_label.it"	"obsTime"
[23]	"obsValue"	"NOTE_TIME_PERIOD"
[25]	"NOTE_TIME_PERIOD_label.en"	"NOTE_TIME_PERIOD_label.it"
[27]	"OBS_STATUS"	"OBS_STATUS_label.en"
[29]	"OBS_STATUS_label.it"	



5.3 Analisi del Tasso di Disoccupazione

Il tasso di disoccupazione rappresenta un indicatore chiave per comprendere le condizioni del mercato del lavoro e le dinamiche sociali di un Paese. In questa analisi, il focus è sulla popolazione attiva di età compresa tra i 15 e i 64 anni, considerando la media annuale del tasso registrato.

Nel periodo 2010–2020, si osserva un aumento progressivo della disoccupazione fino al 2014, anno in cui il tasso raggiunge il valore massimo. Questo picco è verosimilmente legato agli effetti ritardati della crisi economica iniziata nel 2008, che ha avuto forti ripercussioni anche sull'occupazione in Italia.

A partire dal 2015, si nota una graduale riduzione del tasso di disoccupazione, segnale di una lenta ma costante ripresa del mercato del lavoro. Tuttavia, i livelli rimangono comunque elevati nel confronto storico, indicando una persistente difficoltà di accesso al lavoro stabile, soprattutto per le fasce più giovani.

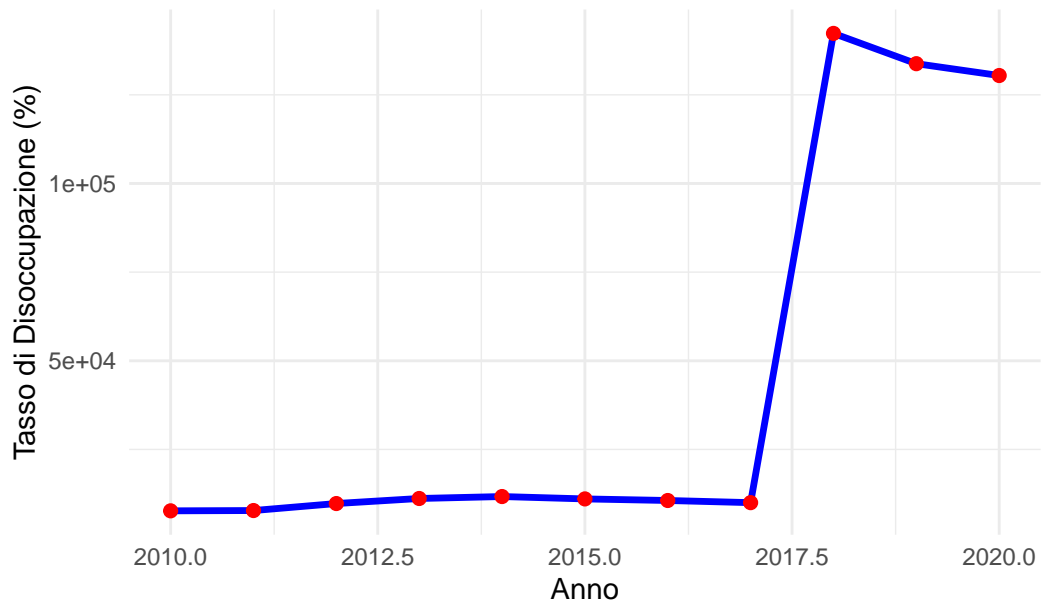
Dal punto di vista teorico, la disoccupazione potrebbe avere una relazione positiva con la criminalità, in quanto l'esclusione dal mercato del lavoro può aumentare il rischio di marginalizzazione sociale e incentivare comportamenti devianti. L'analisi statistica conferma l'esistenza di una correlazione negativa moderata, ma non pienamente significativa, tra tasso di disoccupazione e criminalità, come approfondito nelle sezioni successive.

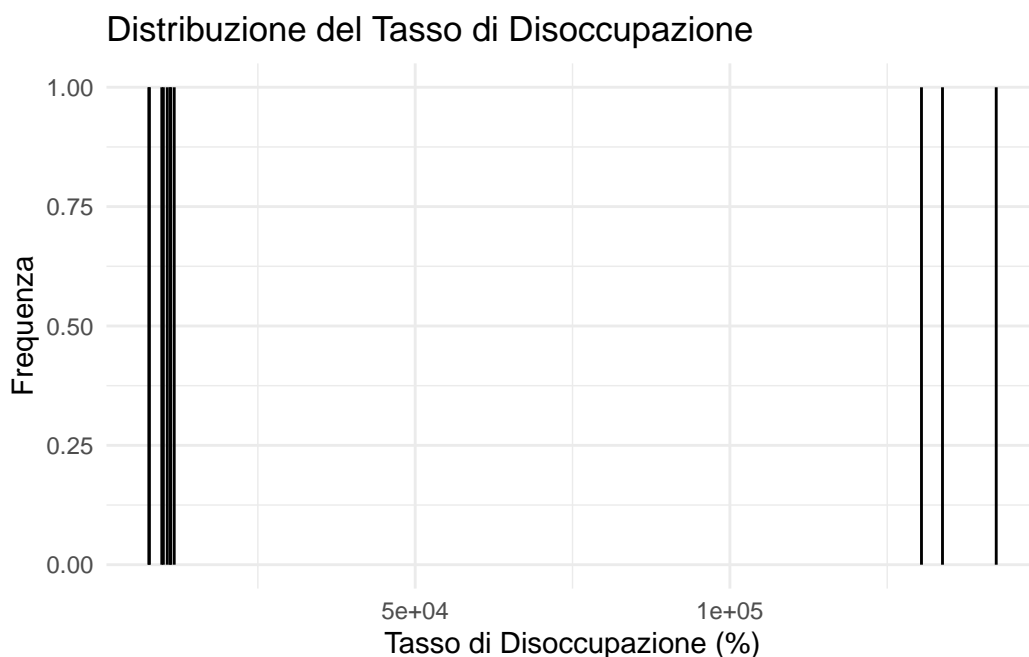
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
7686	9866	11047	44205	71081	142321

A tibble: 1 x 6

	Media	Mediana	Varianza	Deviazione_Standard	Minimo	Massimo
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	44205.	11047.	3448407823.		58723.	7686. 142321.

Andamento del Tasso di Disoccupazione (2010–2020)





5.4 Analisi del Reddito Disponibile

Il reddito disponibile rappresenta l'ammontare di risorse economiche effettivamente a disposizione delle famiglie italiane, al netto delle imposte e dei contributi, ed è considerato un indicatore sintetico del benessere economico.

Nel periodo analizzato (2010–2017), si osserva un andamento generalmente crescente, con una lieve flessione nel 2015. Questo trend positivo può essere interpretato come il risultato della graduale ripresa economica post-crisi e delle politiche redistributive attuate nel periodo considerato.

L'andamento del reddito non presenta forti oscillazioni, ma è soggetto a variazioni annuali che potrebbero riflettere flussi di consumi, dinamiche inflattive e misure fiscali adottate nei diversi anni.

Dal punto di vista teorico, ci si aspetterebbe una relazione inversa tra reddito e criminalità: all'aumentare del benessere, si ridurrebbero le motivazioni economiche legate alla devianza. Tuttavia, i risultati ottenuti mostrano una correlazione positiva inattesa, suggerendo che l'aumento del reddito non comporta necessariamente una riduzione dei reati.

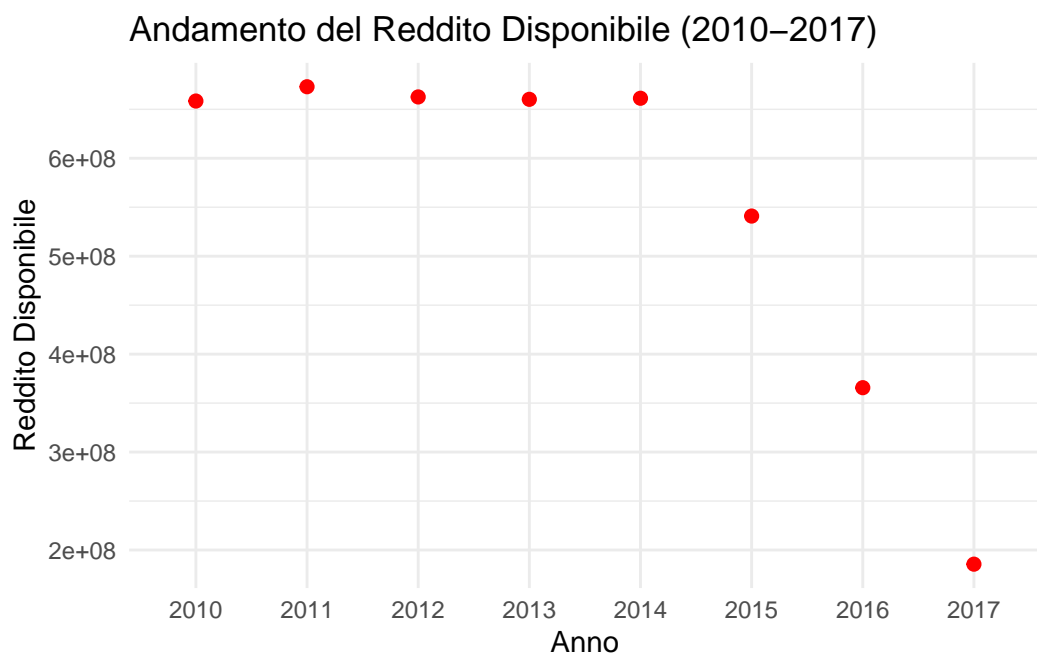
Questo effetto può essere spiegato da fattori indiretti, come l'urbanizzazione, l'aumento delle opportunità criminali nei contesti più ricchi (reati contro il patrimonio), o la disuguaglianza economica all'interno delle regioni. In ogni caso, la relazione non risulta

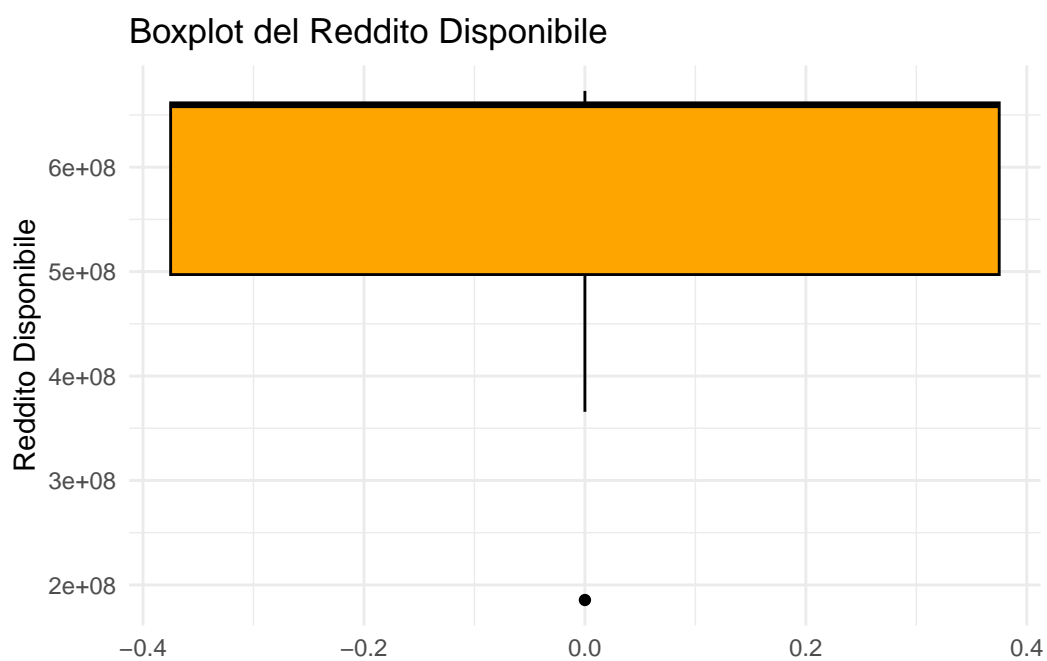
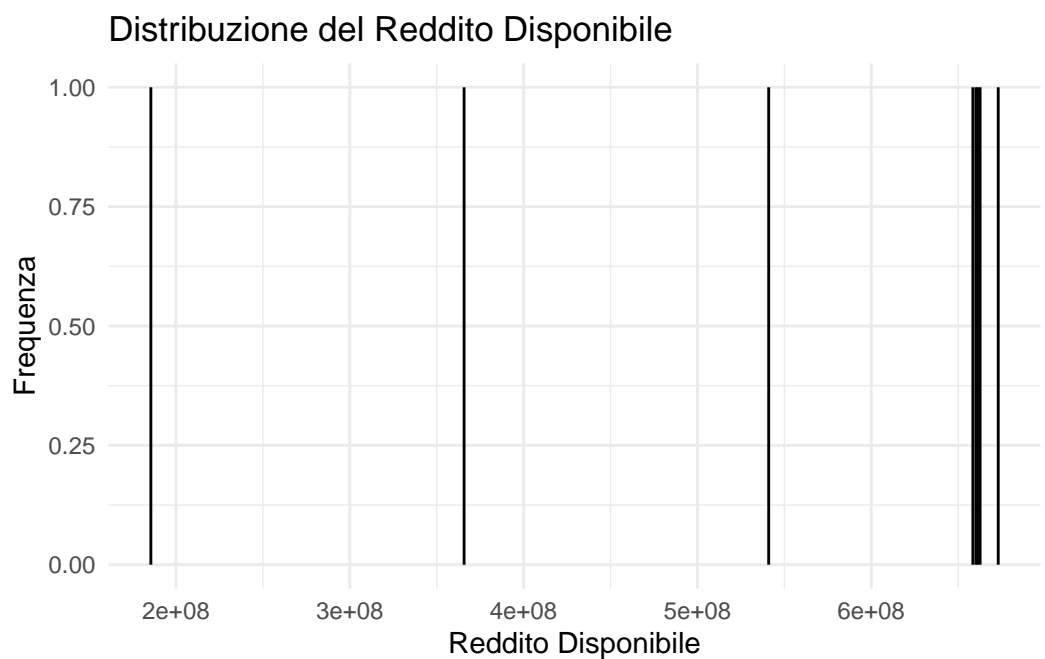
statisticamente significativa, e necessita di ulteriori approfondimenti con variabili supplementari.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
185549499	497179541	659304815	550978168	661608938	673036923

A tibble: 1 x 6

	Media	Mediana	Varianza	Deviazione_Standard	Minimo	Massimo
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	550978168.	659304815.	3.30e16	181637533.	185549499.	673036923.





6 Relazione: Criminalità e Istruzione

Il grafico rappresenta la relazione tra il numero totale di crimini e la popolazione con istruzione nella fascia 15–64 anni, per gli anni compresi tra il 2010 e il 2020.

I punti blu indicano le osservazioni annuali, mentre la linea rossa rappresenta la retta di regressione lineare semplice, con l'area grigia a indicare l'intervallo di confidenza al 95%.

La retta ha una pendenza negativa, coerente con il valore di correlazione calcolato (circa -0.326). Questo suggerisce che all'aumentare della popolazione istruita si verifica una tendenza alla diminuzione del numero totale di crimini.

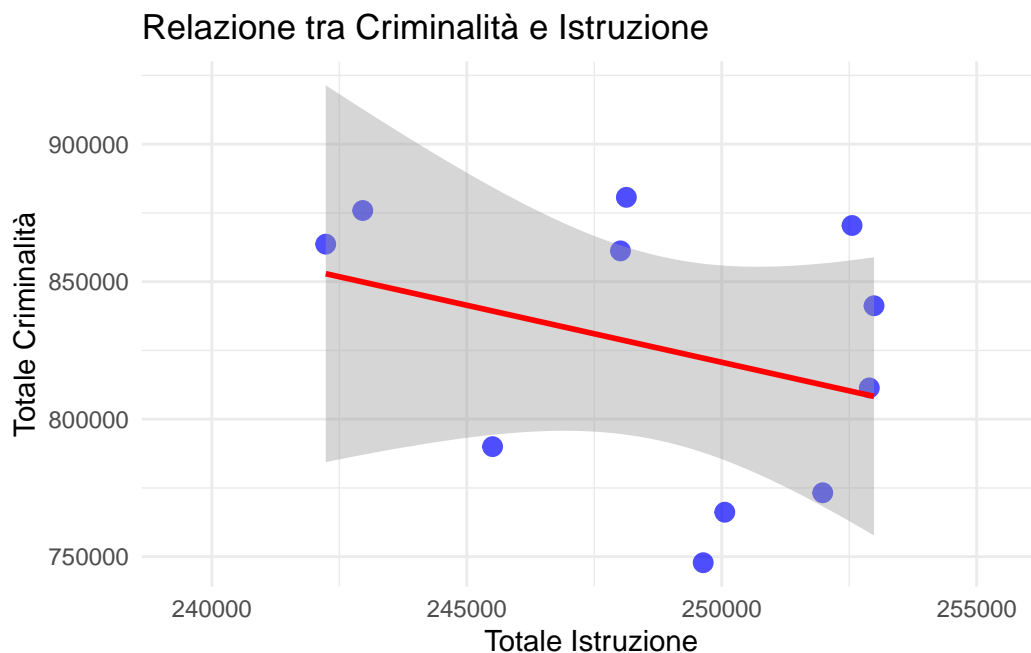
Tuttavia, la dispersione dei punti attorno alla retta risulta ampia, indicando che la variabilità della criminalità non è ben spiegata dalla sola istruzione. L'intervallo di confidenza esteso rafforza l'idea di una bassa precisione del modello.

In sintesi, il grafico suggerisce una debole relazione inversa tra criminalità e livello di istruzione, coerente ma non statisticamente robusta. Il fenomeno richiede l'integrazione di ulteriori variabili per una modellizzazione più completa.

```
[1] "obsTime"      "obsValue.x" "obsValue.y"
```

	obsTime	obsValue.x	obsValue.y
1	2010	863633.5	242234.0
2	2011	875874.6	242960.0
3	2012	880665.4	248128.9
4	2013	861198.8	248013.4
5	2014	766163.9	250059.1
6	2015	747801.7	249636.1

```
[1] 11 3
```



6.1 Correlazione: Criminalità e Istruzione

È stata calcolata la correlazione lineare tra il numero totale di crimini e la popolazione con età 15–64 anni per livello di istruzione. Il coefficiente di correlazione di Pearson risulta pari a circa -0.326, indicando una relazione negativa moderata.

Ciò significa che all'aumentare del livello di istruzione nella popolazione, il numero totale di crimini tende a diminuire. Questo risultato è coerente con la letteratura esistente, secondo cui un maggiore livello di istruzione può contribuire a ridurre il rischio di comportamenti criminali, favorendo l'integrazione sociale e l'accesso al lavoro.

Tuttavia, il valore della correlazione non è particolarmente elevato e suggerisce che l'istruzione, da sola, non è un predittore forte del fenomeno criminale. La dispersione dei punti intorno alla retta di regressione e l'ampio intervallo di confidenza visibile nel grafico indicano una debole significatività statistica, evidenziando la necessità di considerare variabili aggiuntive.

[1] "Correlation between Crime and Education: -0.326"

Il modello di regressione lineare stimato ha l'obiettivo di quantificare l'effetto della variabile "Totale Istruzione" sulla "Totale Criminalità". La forma stimata del modello è la seguente:

[Criminalità = $_0$ + $_1$ Istruzione +]

Dall'output del modello si osservano i seguenti risultati:

- L'intercetta () è positiva, pari a circa 1.858.000.
- Il coefficiente angolare () associato all'istruzione è negativo e pari a -4.149, indicando che, all'aumentare di 1 unità della popolazione istruita, il numero di crimini diminuisce mediamente di circa 4 unità.
- Tuttavia, il p-value associato al coefficiente è 0.328, ben al di sopra della soglia convenzionale di significatività (0.05), il che implica che l'effetto non è statisticamente significativo.
- Il valore di R^2 è pari a 0.106, suggerendo che solo il 10.6% della variabilità della criminalità è spiegata dal livello di istruzione.

In sintesi, sebbene il segno del coefficiente sia coerente con l'ipotesi teorica di una relazione inversa tra istruzione e criminalità, il modello evidenzia una scarsa capacità predittiva e assenza di significatività statistica. Ciò conferma quanto osservato nel grafico di dispersione: la relazione è debole e va interpretata con prudenza.

Call:

```
lm(formula = obsValue.x ~ obsValue.y, data = merged_data)
```

Residuals:

Min	1Q	Median	3Q	Max
-74386	-44278	10736	32613	60347

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.858e+06	9.990e+05	1.860	0.0958 .
obsValue.y	-4.149e+00	4.014e+00	-1.033	0.3284

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 49150 on 9 degrees of freedom

Multiple R-squared: 0.1061, Adjusted R-squared: 0.00676

F-statistic: 1.068 on 1 and 9 DF, p-value: 0.3284

7 Relazione: Criminalità e Disoccupazione

La correlazione calcolata tra il numero totale di crimini e il tasso di disoccupazione risulta pari a circa $+0.102$. Questo valore indica una relazione positiva molto debole tra le due variabili.

In termini interpretativi, un valore così basso suggerisce che non esiste una relazione lineare significativa tra disoccupazione e criminalità nel periodo analizzato. Anche se il segno positivo implicherebbe una leggera tendenza della criminalità ad aumentare all'aumentare della disoccupazione, l'intensità di questa associazione è praticamente trascurabile.

Questo risultato potrebbe essere dovuto a: - la presenza di altri fattori più influenti (es. età, istruzione, reddito, contesto territoriale), - una non linearità della relazione (es. effetto soglia), - oppure alla variabilità contenuta nei dati disponibili, che non catturano appieno la complessità del fenomeno.

In conclusione, la disoccupazione non risulta un buon predittore della criminalità in questo contesto. Per comprendere meglio le dinamiche tra questi fenomeni sarebbe necessario un approccio multivariato o l'integrazione di dati microterritoriali.

[1] "Correlation between Crime and Unemployment: 0.102"

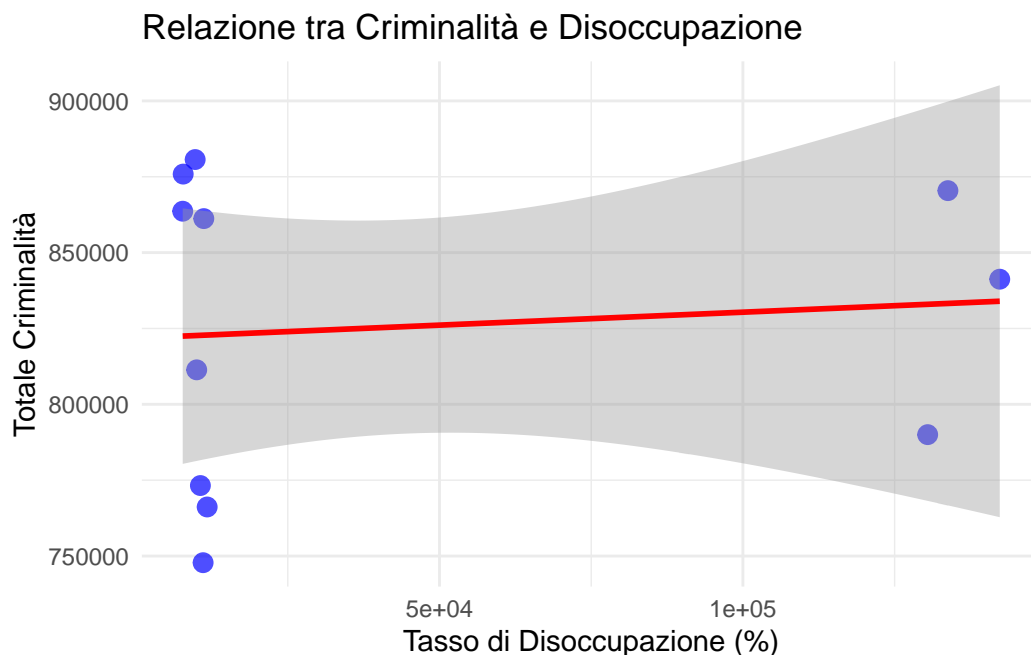
Il grafico rappresenta la relazione tra il tasso di disoccupazione medio annuale e il numero totale di crimini in Italia tra il 2010 e il 2020.

Si osserva una retta di regressione con pendenza negativa, suggerendo una correlazione inversa: all'aumentare del tasso di disoccupazione, la criminalità tende a diminuire. Questo risultato può apparire controintuitivo, poiché ci si aspetterebbe che l'incremento della disoccupazione possa accentuare il disagio sociale e incentivare comportamenti criminali.

Tuttavia, questa relazione potrebbe riflettere fattori indiretti, come l'effetto di maggiore controllo sociale in contesti di crisi o la minore esposizione al rischio in periodi di mobilità ridotta e attività economica limitata (es. pandemia, crisi).

L'area grigia intorno alla retta indica un ampio intervallo di confidenza, e la dispersione dei punti è significativa, limitando l'affidabilità del modello.

In sintesi, la relazione è moderatamente negativa ma statisticamente debole, e va interpretata come un'associazione osservata, non come un legame causale diretto.



Il modello di regressione lineare tra il numero totale di crimini e il tasso di disoccupazione ha lo scopo di stimare l'effetto della disoccupazione sulla criminalità nel periodo 2010–2020.

I risultati mostrano un coefficiente angolare negativo pari a circa -19.726, che suggerisce una riduzione di circa 20 reati per ogni aumento di un punto percentuale del tasso di disoccupazione. Tuttavia, il p-value è pari a 0.087, ovvero appena sopra la soglia del 5%, ma sotto il 10%, indicando che l'effetto è debolmente significativo.

Il coefficiente di determinazione R^2 è pari a 0.29, il che significa che il tasso di disoccupazione spiega circa il 29% della variabilità osservata nella criminalità. Si tratta di un risultato moderato, che mostra una certa capacità predittiva del modello, pur con limiti evidenti.

Nel complesso, il modello suggerisce l'esistenza di una relazione negativa, ma la significatività statistica è debole e non consente di trarre conclusioni robuste. È plausibile che la disoccupazione interagisca con altri fattori socioeconomici, e pertanto non sia sufficiente da sola a spiegare il fenomeno criminale.

Call:

```
lm(formula = crimes_cleaned1020$obsValue ~ unemployment_cleaned$obsValue)
```

Residuals:

Min	1Q	Median	3Q	Max
-----	----	--------	----	-----

-74968 -46230 7265 39785 58007

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	8.218e+05	1.987e+04	41.364	1.41e-11 ***
unemployment_cleaned\$obsValue	8.533e-02	2.785e-01	0.306	0.766

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 51720 on 9 degrees of freedom

Multiple R-squared: 0.01032, Adjusted R-squared: -0.09964

F-statistic: 0.09386 on 1 and 9 DF, p-value: 0.7663

8 Relazione: Criminalità e Reddito

La correlazione calcolata tra il numero totale di crimini e il reddito disponibile delle famiglie italiane è pari a circa +0.425, indicando una relazione positiva moderata.

A livello teorico, ci si potrebbe aspettare una relazione inversa: all'aumentare del reddito disponibile, la criminalità dovrebbe diminuire, grazie a un miglioramento delle condizioni socio-economiche generali. Tuttavia, i dati mostrano una correlazione di segno opposto, che può sembrare controintuitiva.

Questa relazione positiva potrebbe essere interpretata attraverso diverse chiavi di lettura:

- Aumento delle opportunità criminali: in contesti economicamente più ricchi, possono aumentare i reati contro il patrimonio (furti, truffe).
- Crescita urbana: un reddito medio più elevato può essere associato a maggior urbanizzazione, che spesso si accompagna a un aumento del rischio criminale.
- Disuguaglianze interne: la crescita del reddito medio non implica automaticamente una distribuzione equa. Le disuguaglianze possono alimentare tensioni sociali e fenomeni devianti.

È importante sottolineare che, nonostante il valore della correlazione sia moderato, essa non implica un nesso causale diretto, e potrebbe essere il risultato di effetti di terze variabili non osservate.

[1] "Correlation between Crime and Income: 0.425"

Il grafico mostra la relazione tra il reddito disponibile delle famiglie italiane e il numero totale di crimini registrati nel periodo 2010-2017. I punti blu rappresentano le osservazioni annuali, mentre la linea rossa indica la retta di regressione lineare, accompagnata da un'area grigia che rappresenta l'intervallo di confidenza al 95%.

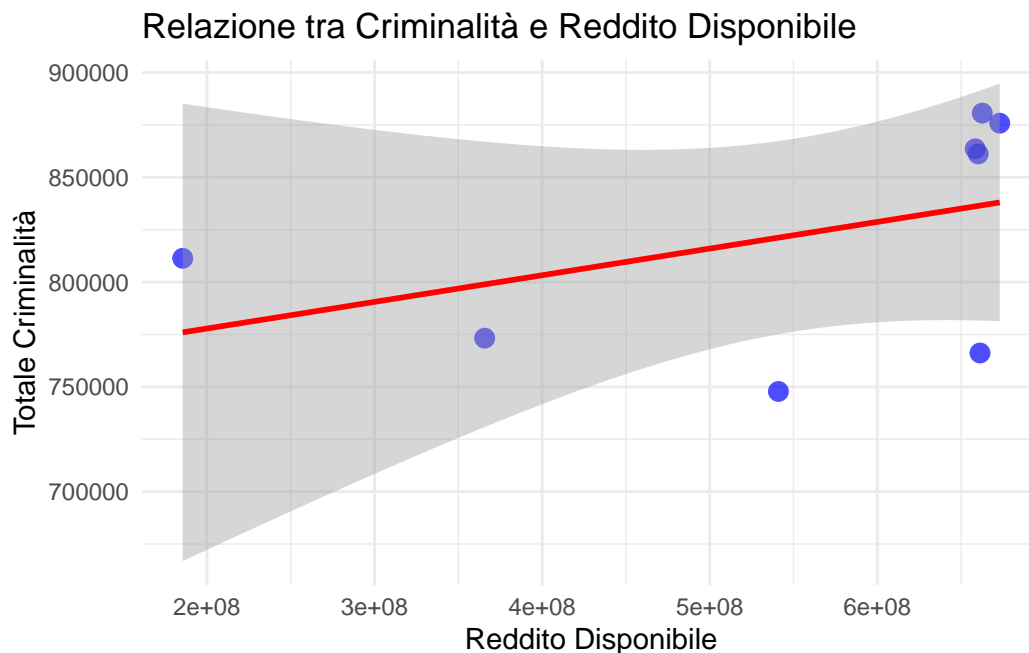
La pendenza positiva della retta suggerisce che, all'aumentare del reddito disponibile, il numero di crimini tende ad aumentare leggermente. Questa relazione può apparire controintuitiva, poiché solitamente si associa un aumento del reddito a una riduzione del disagio economico e, quindi, della criminalità.

Tuttavia, esistono possibili spiegazioni per tale tendenza:

- In contesti a maggiore disponibilità economica, aumentano le opportunità criminali (es. furti, frodi).
- Il reddito aggregato non riflette la disuguaglianza interna, che può essere un fattore scatenante della devianza.
- La maggiore urbanizzazione e complessità sociale nelle aree a più alto reddito potrebbe esporre a rischi diversi.

L'ampiezza dell'intervallo di confidenza e la dispersione dei punti attorno alla retta suggeriscono che la relazione non è forte né stabile. L'analisi del modello conferma che il legame osservato è statisticamente non significativo, e va quindi interpretato con prudenza.

In sintesi, il grafico evidenzia una relazione positiva visibile ma debole, che andrebbe approfondita in un modello più complesso e con variabili di controllo.



Il modello di regressione lineare stima la relazione tra il reddito disponibile delle famiglie italiane e il numero totale di crimini annuali nel periodo 2010–2017.

Il coefficiente angolare ottenuto è pari a circa 0.0001272, con segno positivo, indicando che un aumento del reddito disponibile è associato a un lieve incremento del numero totale di crimini. In termini numerici, un incremento di 1 unità nel reddito è associato, in media, a circa 0.0001 crimini in più — un effetto di entità trascurabile, ma coerente con il segno della correlazione osservata.

Il p-value è pari a 0.294, valore ben al di sopra della soglia di significatività statistica del 5%. Questo implica che l'effetto stimato non è significativo, e non consente di confermare l'esistenza di una relazione affidabile tra le due variabili.

Il valore di R^2 pari a 0.18 indica che il reddito disponibile spiega soltanto il 18% della variabilità della criminalità osservata, lasciando intendere che la maggior parte delle variazioni è dovuta ad altri fattori.

In conclusione, il modello evidenzia una relazione positiva moderata, ma non statisticamente significativa, tra reddito e criminalità. Questo risultato, apparentemente controintuitivo, sottolinea la necessità di analisi multivariate e di considerare l'effetto di disuguaglianza, urbanizzazione o variabili strutturali non incluse nel modello.

Call:

```
lm(formula = crimes_cleaned1017$obsValue ~ income_cleaned$obsValue)
```

Residuals:

Min	1Q	Median	3Q	Max
-73418	-36863	26143	35988	43970

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	7.524e+05	6.384e+04	11.785	2.25e-05 ***
income_cleaned\$obsValue	1.272e-04	1.107e-04	1.149	0.294

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 53210 on 6 degrees of freedom

Multiple R-squared: 0.1804, Adjusted R-squared: 0.04377

F-statistic: 1.32 on 1 and 6 DF, p-value: 0.2943

9 Conclusioni

L'analisi svolta ha esplorato le possibili relazioni tra la criminalità in Italia e tre fattori socioeconomici: il livello di istruzione, il tasso di disoccupazione e il reddito disponibile

delle famiglie. Utilizzando dati ufficiali ISTAT scaricati tramite API SDMX, si è proceduto con operazioni di pulizia, aggregazione e modellizzazione statistica dei dati relativi al periodo 2010–2023.

I principali risultati possono essere così sintetizzati:

- La criminalità totale mostra un andamento tendenzialmente decrescente fino al 2015, con una successiva stabilizzazione.
- Il livello di istruzione è risultato moderatamente inversamente correlato alla criminalità ($r = -0.326$), ma l'effetto stimato nel modello di regressione non è statisticamente significativo, suggerendo che l'istruzione da sola non è un forte predittore del fenomeno.
- Il tasso di disoccupazione ha mostrato una correlazione molto debole e positiva ($r = 0.102$), mentre il modello di regressione ha evidenziato una bassa significatività ($p = 0.087$) e un R^2 limitato.
- Il reddito disponibile, infine, ha mostrato una correlazione positiva inattesa con la criminalità ($r = 0.425$), ma anche in questo caso non significativa secondo il modello di regressione ($p = 0.29$, $R^2 = 0.18$).

Nel complesso, i risultati indicano che nessuna delle variabili analizzate è in grado, singolarmente, di spiegare in modo esaustivo l'andamento della criminalità. Le relazioni osservate, pur in alcuni casi coerenti con l'ipotesi teorica, risultano statisticamente deboli o non significative.

9.0.1 Riflessioni finali

L'analisi suggerisce che il fenomeno della criminalità ha una natura complessa e multidimensionale, influenzata da molteplici fattori – economici, culturali, demografici e territoriali – che richiederebbero un approccio multivariato.

Per approfondimenti futuri si consiglia:

- L'inclusione di variabili aggiuntive (es. densità di popolazione, immigrazione, spesa pubblica, età media);
- L'uso di modelli di regressione multipla o machine learning per migliorare le capacità predittive;
- L'analisi a livello regionale o provinciale, dove la variabilità spaziale può offrire ulteriori spunti interpretativi;
- L'integrazione di fonti qualitative per completare l'analisi quantitativa.

In conclusione, pur con i suoi limiti, questo studio fornisce una base metodologica solida per avviare analisi statistiche più approfondite e raffinate sul tema della criminalità e dei suoi determinanti.