

## *Indice*

1. Il progetto.....	p. 03
a. Il problema	
b. L'obbiettivo	
c. Lo studio delle ipotesi	
2. L'analisi.....	p. 07
a. Il piano di campionamento	
b. Lo studio delle variabili	
c. Il modello di regressione	
3. I risultati e le conclusioni.....	p. 21
4. Riferimenti.....	p. 25
a. Allegati	
b. Bibliografia	
c. Sitografia	
d. Tesi consultate	
e. Fonti normative	

## *Il progetto*

**Sommario:** 1. Il problema; 2. L'obiettivo; 3. Lo studio delle ipotesi.

Il fenomeno del sovraffollamento carcerario è un problema che il nostro Paese ha dovuto da sempre affrontare. Si tratta di una questione che incide in primo luogo sulla gestione ed organizzazione degli istituti penitenziari e, in secondo luogo, sulla sicurezza dei detenuti stessi e delle persone che vi lavorano.

Negli ultimi anni sono stati diversi i provvedimenti che i vari governi hanno adottato per rispondere all'esigenza del sistema penitenziario di alleviare l'eccessiva saturazione creatasi, adottando forme di indulto, depenalizzazione e ampliando leggermente la capacità contenitiva degli istituti carcerari italiani. I risultati hanno mostrato sì qualche miglioramento ma i problemi strutturali di qualche decennio fa tuttora persistono.

Da qui la domanda che ha dato origine a questo lavoro: sarebbe possibile costruire uno strumento capace di analizzare la situazione italiana e di prevedere, con un margine di errore accettabile, quale sarà il numero di detenuti per il periodo selezionato?

L'idea di fondo sarebbe quella di, attraverso qualche concetto di statistica, impostare un modello di regressione il quale, una volta inserite come variabili indipendenti alcuni dati pubblicamente reperibili circa le condizioni economico-sociali del nostro paese, dovrebbe essere in grado di stimare un valore assoluto circa il numero di detenuti effettivi per un certo anno, per un certo carcere.

Vediamo come.

## Il problema

Secondo gli ultimi dati del Ministero della Giustizia<sup>1</sup> la situazione attuale (si veda la figura 1) è la seguente: sul suolo italiano si contano circa 200<sup>2</sup> carceri - tra case circondariali, case di reclusione, istituti a custodia attenuata, residenze per l'esecuzione delle misure di sicurezza, colonie agricole e case di lavoro - per un ammontare di quasi 50'000 posti letto che però non bastano per gli oltre 60'000 detenuti. Il tasso di affollamento medio stimato da Antigone<sup>3</sup> è pari al 120%, dunque 1,2 detenuti per posto letto regolamentare. Il dato potrebbe sembrare non allarmante, ma occorre considerare che la moda della capienza delle carceri italiane, così come si vedrà anche in seguito, risulta essere inferiore ai 200 posti, pertanto anche una minima percentuale di sovraffollamento potrebbe causare non pochi disagi al management penitenziario.

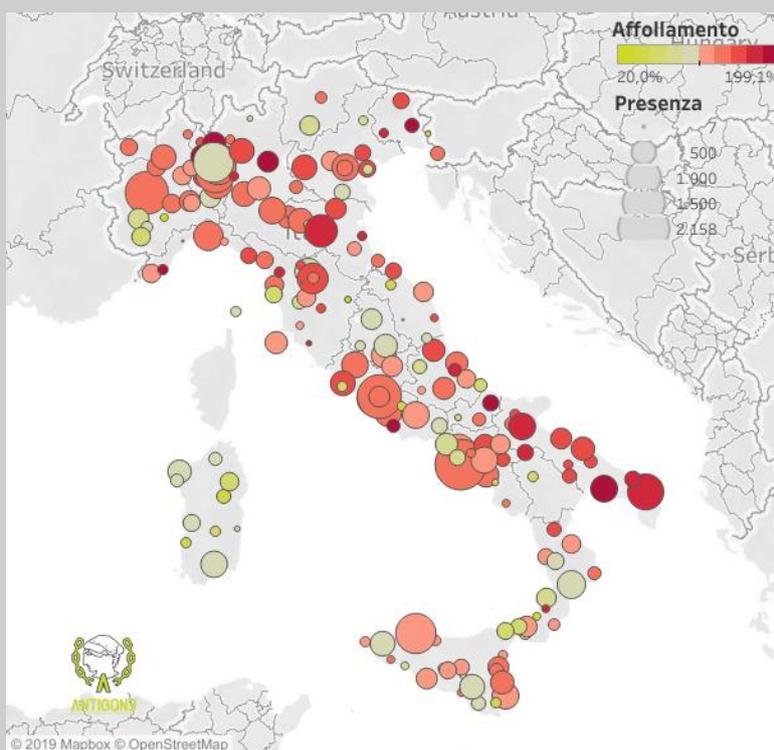


Figura 1 – Situazione italiana al 2019 (fonte: Antigone)

<sup>1</sup> [www.giustizia.it](http://www.giustizia.it), sezione *statistiche*, 2019.

<sup>2</sup> Il dato non considera i singoli edifici penitenziari, ma tutte quelle carceri che dal punto di vista amministrativo sono considerate autonome e dunque dotate di una propria direzione. Annualmente il dato potrebbe subire modifiche per via di correzioni all'assetto amministrativo del sistema penitenziario.

<sup>3</sup> XV rapporto sul carcere dell'Associazione Antigone, [www.antigone.it](http://www.antigone.it) (2019).

### *L'obbiettivo*

Il progetto che sarà qui esposto ha lo scopo di analizzare l'affollamento delle prigioni attraverso lo studio di diverse ipotesi che consentano di stimare un modello di regressione che, data una serie di variabili indipendenti, restituisca una stima dell'affollamento - e relativo sovraffollamento o sottoaffollamento - delle carceri, evidenziando in che misura ciascuna variabile sia in grado di spiegare il fenomeno.

Tale modello potrebbe suggerire al policy maker dove sia necessario agire per correggere il fenomeno e potrebbe agevolare il Dipartimento dell'Amministrazione Penitenziaria, oltre che ciascuna direzione carceraria, nella gestione dell'incontro tra "domanda" e "offerta" di detenzione.

### *Lo studio delle ipotesi*

Quali potrebbero essere le questioni che influiscono sull'affollamento penitenziario? È stata questa la domanda da cui sono scaturite diverse riflessioni in merito al fenomeno in analisi e che hanno portato alla definizione di un'ipotesi di partenza: una maggiore presenza di detenuti potrebbe essere spiegata da condizioni culturali, economiche, sociali e climatiche sfavorevoli. Da qui sono sorti diversi interrogativi sulla definizione stessa di tali condizioni e su come poterle quantificare al fine di considerarle all'interno di un modello statistico. Al termine di un ampio ragionamento sulla questione è stata identificata una serie di oltre 50 variabili semplici<sup>4</sup>, analizzate inizialmente da un punto di vista storico e successivamente da un punto di vista utilitaristico relativamente agli scopi del progetto.

Le variabili semplici considerate spaziano dai numeri relativi agli stessi istituti penitenziari (presenze, capienze, affollamento, ...) a tutti quegli elementi connessi alle condizioni sociali, culturali, economiche e climatiche della popolazione (reddito medio, reati denunciati, livelli d'istruzione,

---

<sup>4</sup> Si veda l'allegato A per il dettaglio delle variabili considerate e delle relative fonti. Nello stesso allegato sono elencate anche le variabili complesse costruite attraverso lo studio di quelle semplici.

numerosità delle famiglie, indici climatici, ...). Ben consci che un numero così elevato di variabili semplici avrebbe potuto rendere macchinosa la costruzione, e la successiva analisi, del modello si è provveduto ad ipotizzare degli aggregati e dunque delle variabili complesse.

## *L'analisi*

**Sommario:** 1. Il piano di campionamento; 2. Lo studio delle variabili; 3. Il modello di regressione.

### *Il piano di campionamento*

Sul territorio italiano insistono diverse tipologie di strutture penitenziarie: case circondariali, case di reclusione, residenze per l'esecuzione delle misure di sicurezza, istituti a custodia attenuata, case di lavoro e colonie agricole. Queste si differenziano tra loro per la tipologia di detenuti che sono designate ad ospitare. Per semplicità in questo studio sono state considerate esclusivamente le case circondariali (CC) e le case di reclusione (CR) miste, poiché si tratta della tipologia maggiormente diffusa. Dagli oltre 200 istituti di partenza, la popolazione statistica considerata ammonta quindi a 184 unità organizzate in 11 provveditorati<sup>5</sup>. La distribuzione delle unità statistiche per dimensioni e localizzazione geografica non è omogenea: in alcuni provveditorati insistono pochi istituti e di dimensioni ridotte, in altri invece ve ne sono di più e con dimensioni variabili.

Considerando quanto detto finora risulta sconveniente l'utilizzo di una tecnica di campionamento diversa da quella stratificata. Infatti, si è da subito provveduto ad organizzare la popolazione statistica in quattro strati basati sulla capienza regolamentare di ciascun istituto penitenziario. La tabella che segue illustra gli strati definiti.

I strato	da 1 a 200 posti letto
II strato	da 201 a 500 posti letto
III strato	da 501 a 900 posti letto
IV strato	oltre 900 posti letto

---

<sup>5</sup> Unità periferiche del Ministero della Giustizia distribuite per aree macro-regionali. L'elenco è reperibile al sito [https://www.giustizia.it/giustizia/it/mg\\_12\\_3\\_9.wp](https://www.giustizia.it/giustizia/it/mg_12_3_9.wp).

Valutando attentamente la composizione della popolazione si è deciso di estrarre un campione di grandezza  $n=60$ , all'interno del quale sono stati inclusi tutti e 6 gli istituti appartenenti al quarto strato. Dopo aver strutturato la popolazione secondo un'allocazione proporzionale per provveditorato secondo i primi tre strati, la restante parte del campione<sup>6</sup> è stata estratta attraverso un campionamento casuale semplice senza ripetizione.

### *Lo studio delle variabili*

In considerazione delle ipotesi descritte in precedenza e come già accennato, sono state definite ed analizzate oltre 50 variabili semplici da alcune delle quali è stato possibile creare più di 10 variabili complesse. Come verrà illustrato successivamente, nella sezione che riguarda i modelli di regressione sviluppati, non tutte le variabili sono state utilizzate per gli scopi del progetto dal momento che i test di correlazione con il numero di detenuti effettivi di ogni istituto penitenziario campionato (variabile di risposta) hanno restituito un esito negativo. Attraverso il software R sono stati eseguiti i test di Pearsons per ciascuna correlazione e da questi, per alcune, è emerso un p-value  $>0.05$  e dunque sono state scartate. Di seguito si riporta il codice di programmazione in linguaggio R utilizzato per i test:

```
cor.test(Dataset$variabile_risposta;Dataset$variabileX)
```

Alla luce di quanto detto finora, queste sono le variabili effettivamente utilizzate<sup>7</sup>:

- **Detenuti effettivamente presenti in ciascun carcere.**

Il numero di detenuti effettivi rappresenta l'affollamento delle carceri, dunque è il focus dell'intero progetto e, come vedremo, per la stima dei modelli di regressione è stata considerata quale variabile di risposta. Attraverso lo studio di questa variabile, in combinazione con

---

<sup>6</sup> Per i dettagli sul piano di campionamento e sul campione estratto si veda l'allegato B, il quale contiene la lista di campionamento, i calcoli dell'allocazione proporzionale e la lista campione. L'elaborazione è stata effettuata con il software Excel.

<sup>7</sup> Per le analisi descrittive delle variabili non utilizzate e dunque non commentate nella presente relazione, si faccia riferimento all'allegato C, dove tra le altre cose vi sono anche le analisi complete delle variabili già commentate.

la capienza effettiva dell'istituto oggetto di analisi, è possibile ricavare il tasso (in termini %) di sovraffollamento o sotto-affollamento, consentendo al policy maker - o a chiunque sia interessato - di avere un quadro generale di quella che è “la domanda” e di quella che è “l’offerta” di detenzione in modo da poter adottare le misure necessarie.

Di questa variabile è stata effettuata un’analisi descrittiva storica relativa al periodo compreso tra il 2017 ed il 2019. Da tale analisi è emerso che in media ogni anno sono detenute nelle nostre carceri circa 60'000 persone, di cui circa l’80% distribuito equamente tra Nord e Sud Italia e il restante 20% concentrato nel Centro Italia. Emerge chiaramente che la distribuzione dei detenuti non segue esattamente la distribuzione delle prigioni sul territorio nazionale: in alcune zone dove vi sono più istituti non sempre vi sono molti detenuti. Probabilmente dietro la localizzazione geografica delle carceri non vi è stato, da parte dei progettisti, un ragionamento orientato a distribuire sul territorio edifici con una capienza omogenea.

- **Capienza effettiva degli istituti carcerari.**

Ricollegandoci al discorso poc’anzi concluso, viene spontaneo prendere in esame, dunque, anche la capienza delle carceri italiane della quale è stata fatta un’analisi descrittiva storica relativa al periodo compreso tra il 2017 e il 2019. Dall’istogramma che segue è possibile notare come la metà degli istituti penitenziari italiani non raggiunge i 200 posti letto regolamentari.

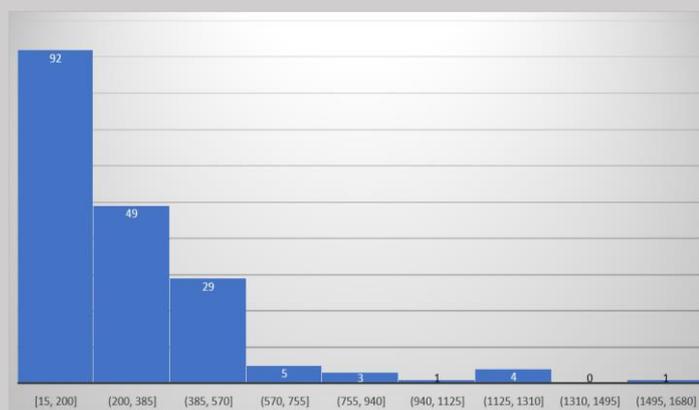


Figura 2 - Distribuzione delle capienze delle prigioni italiane al 2019 (Fonte: nostra elaborazione di dati del D.A.P.)

Ogni anno in Italia sono disponibili circa 50'000 posti letto di cui poco più del 30% nel Nord-Italia, il 45% nel Sud-Italia e il restante 25% nel Centro-Italia.

- **Criminalità.**

Portando avanti uno studio sull'affollamento delle prigioni non può essere trascurata la criminalità, con le relative tendenze su base territoriale: sembra naturale che se in carcere ci finiscono le persone che hanno commesso un reato, qualche relazione tra l'andamento dei reati e l'affollamento carcerario dovrà pur esistere (a prescindere dalla consistenza). Criminalità e carcere s'ipotizza che siano tra loro due questioni strettamente legate e che possono andare in diverse direzioni. Nello studiare la variabile tasso di criminalità è dato per assunto che nell'arco temporale in esame (2010-2018) non siano stati incrementati e nemmeno diminuiti/modificati i fatti previsti dalla legge come reato.

La variabile tasso di criminalità può essere considerata come una variabile complessa, costituita da 21 variabili semplici (diverse tipologie di reati<sup>8</sup>) che insieme forniscono una misura della criminalità in un certo luogo. In linea generale è stata effettuata un'analisi storica descrittiva per il periodo di riferimento, mentre per il solo 2017 (anno di riferimento dell'indagine) è stato costruito l'indice di criminalità ad hoc (sia su base provinciale che regionale).

Il suddetto indice è stato elaborato con due approcci differenti: il primo ha previsto, con l'ausilio del software R, un'organizzazione in 5 insiemi delle 21 tipologie di reato attraverso l'applicazione del metodo gerarchico di analisi per gruppi; a questi ultimi, contenenti fattispecie di reati tra loro omogenei, è stata applicata l'analisi per componenti principali al fine di ottenere un'unica variabile latente (è stata presa la prima dimensione) considerata come indice di criminalità in riferimento a quello specifico gruppo di reati per ciascuna provincia e regione. Il secondo approccio ha visto invece

---

<sup>8</sup> Per l'elenco completo si rimanda all'allegato A.

direttamente l'applicazione dell'analisi per componenti principali su tutte le 21 tipologie al fine d'individuare un'unica variabile latente (la prima dimensione) da utilizzare come indice di criminalità per ciascuna provincia e regione riferita a tutti i reati in generale. Riassumendo: il primo approccio vede quindi 5 variabili complesse, il secondo invece ne vede una sola. In riferimento a quest'ultima il grafico che segue riassume la situazione italiana al 2017:



Figura 3 - Mappa della criminalità al 2017 (Fonte: nostra elaborazione di dati ISTAT)

La figura 3 illustra l'indice di criminalità che varia da un minimo di -8.52 (pochi crimini e spesso di lieve entità) ad un massimo di 19.65 (criminalità maggiormente diffusa e di entità variabile). Come si può notare, le regioni dove il fenomeno presenta valori più alti sono: il Lazio, la Lombardia, la Campania, la Puglia e la Sicilia. Diversamente in Valle d'Aosta, Molise e Basilicata la criminalità non sembra essere un problema.

Di seguito si riporta il codice generico utilizzato in R per l'analisi dei gruppi e per componenti principali.

```

1 library(cluster)
2 Crimine1<-Crimine[,2:22]
3 Crimine1=na.omit(Crimine1)
4 Crimine2<- scale(Crimine1)
5 Crimine3<- hclust(dist(Crimine2),method="ward.D")
6 plot(Crimine3)
7 rect.hclust(Crimine3, 5)
8 Gruppi<-cutree(Crimine3, 5)
9 silhouette<-silhouette(Gruppi,dist(Crimine3),nmax.lab=60)
10 summary(silhouette)
11 plot(silhouette, main="", col=1:3)
12 Crimine3Sil = cbind(Dataset[,1],silhouette[,1:3])
13 Crimine3sil
14 G1<-subset(Crimine3sil,Crimine3Sil$cluster=="1")
15 G2<-subset(Crimine3sil,Crimine3Sil$cluster=="2")
16 G3<-subset(Crimine3sil,Crimine3Sil$cluster=="3")
17 G4<-subset(Crimine3sil,Crimine3Sil$cluster=="4")
18 G5<-subset(Crimine3sil,Crimine3Sil$cluster=="5")
19 G1$Ist
20 G2$Ist
21 G3$Ist
22 G4$Ist
23 G5$Ist

```

Figura 4 - Codice R per analisi dei gruppi

```

1 library(FactoMineR)
2 pca<- PCA(Crimine, scale.unit =TRUE,
3 quali.sup=1)
4 summary(pca)
5 pca$ind$coord
6
7 Crimine$Dim1<-pca$ind$coord[,1]
8 plot(Crimine$Dim1)
9 write.table(Crimine, file="Crimine.csv",
10 quote=F, sep=";", dec=".", na="",
11 row.names=T, col.names=T)

```

Figura 5 - Codice R per analisi in componenti principali

- **Integrazione.**

Relativamente allo studio oggetto della presente relazione non può non suscitare interesse il fatto che tanti detenuti siano stranieri. Non è una questione di razzismo o discriminazione, bensì un semplice dato

statistico: nelle prigioni italiane, in linea generale, un detenuto su tre è straniero. Su circa 60'000 detenuti al giugno 2019, risultano indicativamente 20'000 stranieri. Ciò è rilevante nell'ambito dell'analisi sull'affollamento penitenziario dal momento che non può non esistere una relazione tra questo dato e tutto ciò che ha a che vedere con l'integrazione degli stranieri nel tessuto sociale italiano. L'ipotesi di partenza è che per un'area caratterizzata da un livello d'integrazione basso ci sono più stranieri nelle prigioni rispetto ad aree dove questi ultimi sono meglio integrati.

La variabile integrazione ha una struttura complessa derivante da un'analisi per componenti principali, effettuata sempre con R, delle seguenti variabili semplici per l'anno 2017: a) Presenze degli stranieri, b) Stranieri occupati, c) Stranieri di 6+ anni iscritti ad un regolare corso di studi, d) Tasso di acquisizione della cittadinanza italiana, e) Sentimento d'italianità: quanti studenti immigrati di seconda generazione hanno dichiarato di sentirsi italiani.

Dell'analisi per componenti principali, effettuata soltanto per l'anno 2017, sono state considerate le prime due dimensioni: integrazione 1 (la situazione degli stranieri in Italia dal punto di vista del lavoro e della formazione) e integrazione 2 (la situazione degli stranieri dal punto di vista sociale). È necessario segnalare che le variabili semplici poc'anzi descritte sono state considerate relativamente al periodo 2011-2018, per il quale è stata riscontrata una carenza di dati al dettaglio provinciale ed è stato ritenuto opportuno procedere ad una ricostruzione dei dati mancanti. A tal scopo si è utilizzata un'analisi per regressione<sup>9</sup> basata sui dati disponibili (dettaglio provinciale per il 2011 e dettaglio regionale per il periodo 2011-2018). In linea generale per il periodo di riferimento è stata effettuata anche un'analisi descrittiva storica dalla quale si evince che in alcune province (quelle a vocazione industriale) gli stranieri godono di una situazione economica più stabile seppure il sentimento d'italianità non è eccessivamente diffuso rispetto invece ad alcune

---

<sup>9</sup> I dettagli tecnici della ricostruzione sono disponibili nell'allegato C (Integrazione.pdf).



condizioni culturali 2 (livello d'istruzione) agganciate ad una singola provincia per l'anno 2017. Invece per il periodo 2015-2018 è stata effettuata un'analisi descrittiva delle singole variabili semplici e da questa si evince che nel centro-nord vi è una maggiore concentrazione di persone disposte a partecipare alla vita culturale e con un livello d'istruzione tendenzialmente più alto, mentre nel mezzogiorno si riscontra una situazione più o meno simmetrica con alcune eccezioni quali ad esempio la provincia di Napoli. L'ipotesi per la quale è stata considerata la variabile delle condizioni culturali vuole che all'aumentare della positività di queste diminuisca il numero di detenuti presenti nelle carceri.

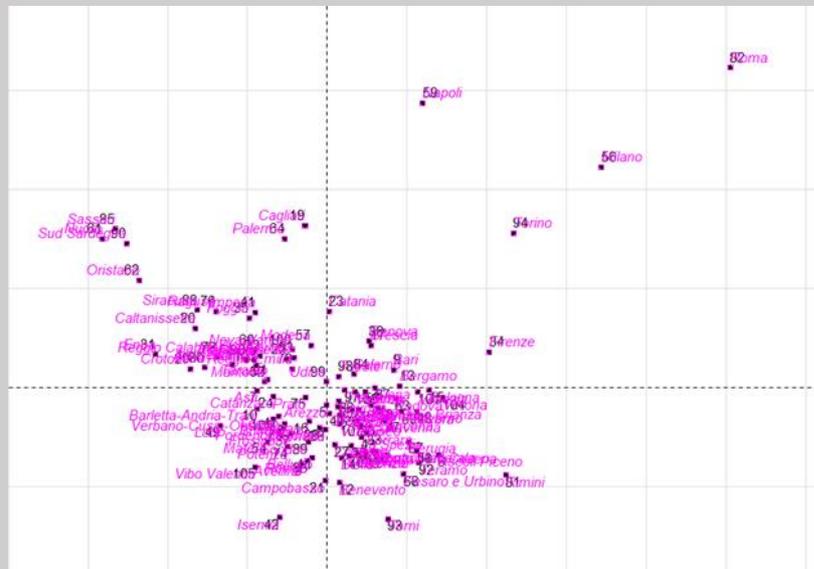


Figura 8 - Posizionamento delle province italiane sulla base delle condizioni culturali (Fonte: nostra elaborazione di dati ISTAT e Sole 24 Ore)

```

1 library(FactoMineR)
2 pca<- PCA(Cculturali, scale.unit =TRUE,
3 quali.sup=1)
4 summary(pca)
5 pca$ind$coord
6
7 Cculturali$Dim1<-pca$ind$coord[,1]
8 Cculturali$Dim2<-pca$ind$coord[,2]
9 plot(Cculturali$Dim1)
10 plot(Cculturali$Dim2)
11 write.table(Cculturali, file="Cculturali.csv",
12 quote=F, sep=";", dec=".", na="", row.names=T,
13 col.names=T)

```

Figura 9 - Codice R per l'analisi in componenti principali

- **Area macro-regionale.**

Questa variabile è stata presa in considerazione dal momento che si è ipotizzata una possibile influenza della localizzazione geografica degli istituti penitenziari sul numero effettivo di detenuti. Semplificando: il fatto che un carcere si trovi a nord piuttosto che al sud, può influire in qualche modo sull'affollamento? Chiaramente un'analisi di questa variabile a livello descrittivo sarebbe tecnicamente impossibile, infatti non è stata elaborata. È stato comunque ritenuto opportuno inserire la variabile nel modello di regressione, come si vedrà, al fine di verificarne la significatività relativamente al fenomeno in esame. È stato possibile fare questa operazione trasformando, con l'ausilio di R, "Area macro-regionale" in 3 dummies: "Nord-est", "Nord-ovest" e "Mezzogiorno" assumendo "Centro" come reference-category. Di seguito il codice utilizzato per la trasformazione.

```
Dataset$Area macro-regionale=as.factor(Dataset$`Area macro-regionale`)
```

*Figura 10 - Codice R per la trasformazione in Dummy*

### *Il modello di regressione*

Considerando le variabili fin qui esposte, in virtù di quanto detto a proposito dei test di correlazione, è stato costruito con l'ausilio di R un primo modello di regressione basato sulle capienze degli istituti penitenziari campionati, sui valori dell'indice d'integrazione 1 e delle condizioni culturali a questi ultimi associati sulla base della provincia dove sono localizzati e sul valore dell'indice della criminalità associato al gruppo di reati n.5 (rapine, contraffazioni) per le stesse province.

```

1 Call:
2 lm(formula = A$Deteff ~ A$Cap + A$G5 + A$Int1 + A$Condc1 + A$Condc2)
3
4 Residuals:
5      Min       1Q   Median       3Q      Max
6 -290.890  -28.929   4.094   33.228  194.432
7
8 Coefficients:
9             Estimate Std. Error t value Pr(>|t|)
10 (Intercept)   7.79638    17.52543   0.445  0.65820
11 A$Cap         1.12765     0.04806  23.464 < 2e-16 ***
12 A$G5         26.32296     9.14673   2.878  0.00572 **
13 A$Int1        0.11297     6.55254   0.017  0.98631
14 A$Condc1     -5.07040     9.06142  -0.560  0.57809
15 A$Condc2    -11.08908    13.34021  -0.831  0.40949
16 ---
17 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
18
19 Residual standard error: 86.16 on 54 degrees of freedom
20 Multiple R-squared:  0.9652, Adjusted R-squared:  0.962
21 F-statistic: 300 on 5 and 54 DF, p-value: < 2.2e-16

```

Figura 11 - Summary modello di regressione #1

Questo primo modello può essere sintetizzato dalla seguente funzione:

*Detenuti\_presenti*

$$\begin{aligned}
&= 7.79638 + (\text{capienza\_carceri} * 1.12765) + (\text{Gruppo\_criminalità5} \\
&* 26.32296) + (\text{Integrazione1} * 0.11297) + (\text{Condizioni\_culturali1} \\
&* -5.0704) + (\text{Condizioni\_culturali2} * -11.08908) + \mathcal{E}
\end{aligned}$$

Come si evince dal summary, soltanto “Capienza\_carceri” e “Gruppo\_criminalità5” risultano significative. Nel complesso il modello risulta significativo con un p-value <2.2e-16 ed un R<sup>2</sup> adjusted pari a 0.962 tale per cui il modello è in grado di spiegare oltre il 96% del fenomeno in esame. Per ogni variazione unitaria dei detenuti presenti (variabile di risposta), la capienza delle carceri subisce un aumento pari a 1.13 e la criminalità relativa ai reati considerati accusa un incremento di 26.3. La variazione delle restanti variabili, essendo non significative, è assimilabile a 0. Il risultato sembrerebbe soddisfacente, ma con dei margini di miglioramento. Al fine di consolidare i risultati relativamente ai coefficienti significativi, sono state rimosse le variabili con un p-value individuale >0.05.

Il modello così ottenuto, che presenta lo stesso R<sup>2</sup> oltre che la stessa significatività, è ben sintetizzato nella seguente funzione:

*Detenuti\_presenti*

$$= 7.79638 + (\text{capienza\_carceri} * 1.12765) + (\text{Gruppo\_criminalità} * 26.3229) + \varepsilon$$

Come precedente accennato, la variabile “Criminalità” è stata costruita secondo due approcci: il primo per gruppi di reati, il secondo considerando i reati nel loro complesso. Si è deciso d’impostare un terzo modello di regressione lasciando invariate le altre variabili e modificando quella della criminalità utilizzando, questa volta, la sua costruzione sulla base del secondo approccio che comunque presenta una significativa correlazione con la variabile di risposta. Di seguito viene riportato il summary dei risultati consolidati per i coefficienti significativi e la relativa funzione.

```

1 Call:
2 lm(formula = A$Deteff ~ A$Cap + A$P)
3
4 Residuals:
5      Min       1Q   Median       3Q      Max
6  -326.71  -29.55    5.47   40.06  194.95
7
8 Coefficients:
9             Estimate Std. Error t value Pr(>|t|)
10 (Intercept)  -1.35824   16.47251  -0.082   0.935
11 A$Cap         1.15746    0.04287  26.998 <2e-16 ***
12 A$P          4.60372    1.97065   2.336  0.023 *
13 ---
14 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
15
16 Residual standard error: 88.29 on 57 degrees of freedom
17 Multiple R-squared:  0.9615, Adjusted R-squared:  0.9601
18 F-statistic: 711.4 on 2 and 57 DF,  p-value: < 2.2e-16

```

Figura 12 - Summary modello di regressione #2

*Detenuti\_presenti*

$$= -1.35824 + (\text{capienza\_carceri} * 1.15746) + (\text{CriminalitàP} * 4.60372) + \varepsilon$$

Anche quest'ultimo modello presenta risultati altrettanto significativi, assimilabili a quelli ottenuti in precedenza. Per ogni variazione unitaria dei detenuti presenti (variabile di risposta), la capienza delle carceri subisce un aumento di 1.15, mentre la criminalità a livello provinciale considerando i reati in generale accusa un incremento di 4.6.

Volendo provare a migliorare i risultati finora ottenuti, è stato costruito un quarto modello di regressione del quale si riportano qui il summary e la funzione.

```

1 Call:
2 lm(formula = Mod$`Detenuti effettivi al 2017` ~ Mod$`Capienze effettive al 2017` +
3   Mod$`Criminalità Gruppo 5 al 2017` + Mod$`Integrazione 1 al 2017` +
4   Mod$`Condizioni culturali 1 al 2017` + Mod$`Condizioni culturali 2 al 2017` +
5   Mod$`Area macro-reg al 2017`)
6
7 Residuals:
8     Min       1Q   Median       3Q      Max
9 -260.772  -17.474    2.335   29.737  200.776
10
11 Coefficients:
12
13             Estimate Std. Error t value Pr(>|t|)
14 (Intercept)      -40.38026    35.79954  -1.128  0.26462
15 Mod$`Capienze effettive al 2017`      1.13357     0.04558  24.869 < 2e-16 ***
16 Mod$`Criminalità Gruppo 5 al 2017`    27.52072     9.25050   2.975  0.00447 **
17 Mod$`Integrazione 1 al 2017`      -7.06900     7.48372  -0.945  0.34932
18 Mod$`Condizioni culturali 1 al 2017`  -1.07481    12.32539  -0.087  0.93085
19 Mod$`Condizioni culturali 2 al 2017`  -8.45983    12.84706  -0.659  0.51318
20 Mod$`Area macro-reg al 2017`Mezzogiorno  35.27555    40.77692   0.865  0.39104
21 Mod$`Area macro-reg al 2017`Nord-est   126.37444    43.67588   2.893  0.00559 **
22 Mod$`Area macro-reg al 2017`Nord-ovest  53.82268    42.79326   1.258  0.21422
23 ---
24 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
25
26 Residual standard error: 81.64 on 51 degrees of freedom
27 Multiple R-squared:  0.9705,    Adjusted R-squared:  0.9659
28 F-statistic:  210 on 8 and 51 DF,  p-value: < 2.2e-16

```

Figura 13 - Summary modello di regressione #3

*Detenuti\_presenti\_centro*

$$= -40.38026 + (\text{capienza\_carceri} * 1.13357) + (\text{Criminalità\_gruppo5} * 27.52072) + (\text{Integrazione1} * -7.069) + (\text{Condizioni\_culturali1} * -1.07481) + (\text{Condizioni\_culturali2} * -8.45983) + \mathcal{E}$$

*Detenuti\_presenti\_Mezzogiorno*

$$= -40.38026 + (\text{capienza\_carceri} * 1.13357) + (\text{Criminalità\_gruppo5} * 27.52072) + (\text{Integrazione1} * -7.069) + (\text{Condizioni\_culturali1} * -1.07481) + (\text{Condizioni\_culturali2} * -8.45983) + 35.27555 + \mathcal{E}$$

*Detenuti\_presenti\_Nord – Est*

$$= -40.38026 + (\text{capienza\_carceri} * 1.13357) + (\text{Criminalità\_gruppo5} * 27.52072) + (\text{Integrazione1} * -7.069) + (\text{Condizioni\_culturali1} * -1.07481) + (\text{Condizioni\_culturali2} * -8.45983) + 126.37444 + \mathcal{E}$$

*Detenuti\_presenti\_Nord – Ovest*

$$= -40.38026 + (\text{capienza\_carceri} * 1.13357) + (\text{Criminalità\_gruppo5} * 27.52072) + (\text{Integrazione1} * -7.069) + (\text{Condizioni\_culturali1} * -1.07481) + (\text{Condizioni\_culturali2} * -8.45983) + 53.82268 + \mathcal{E}$$

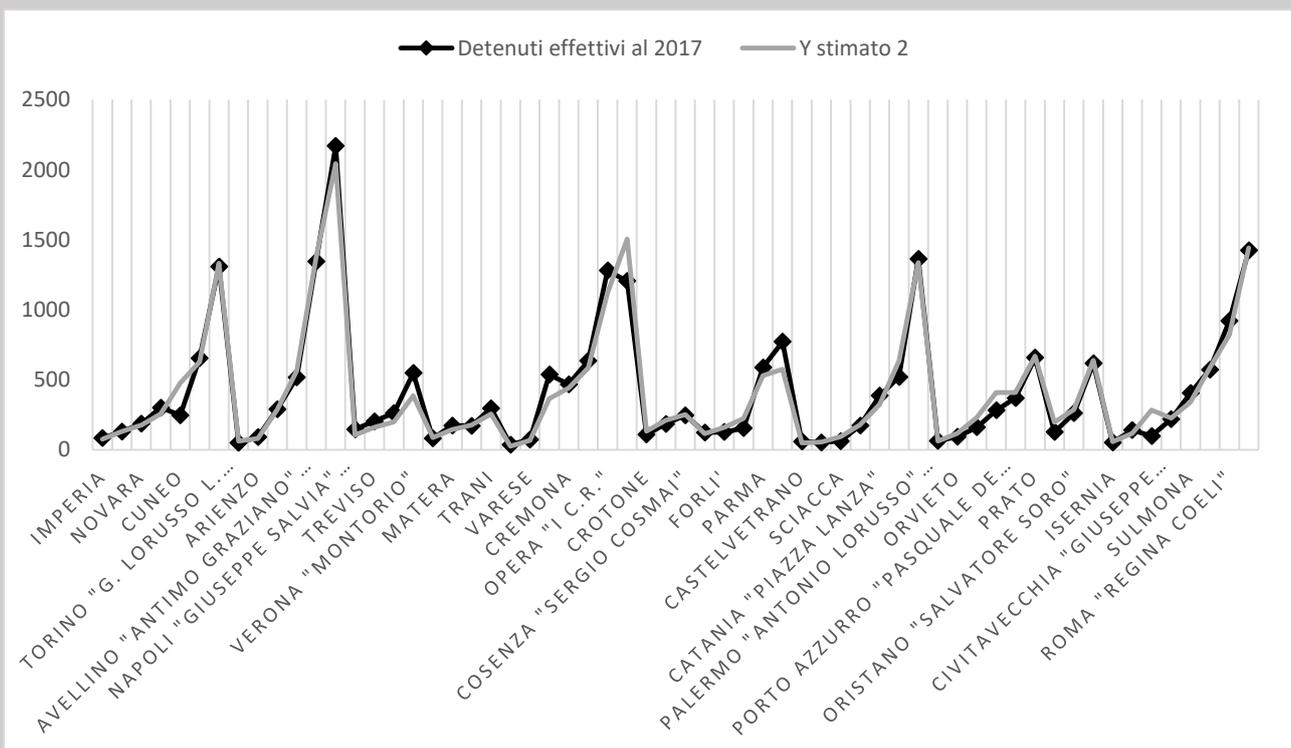
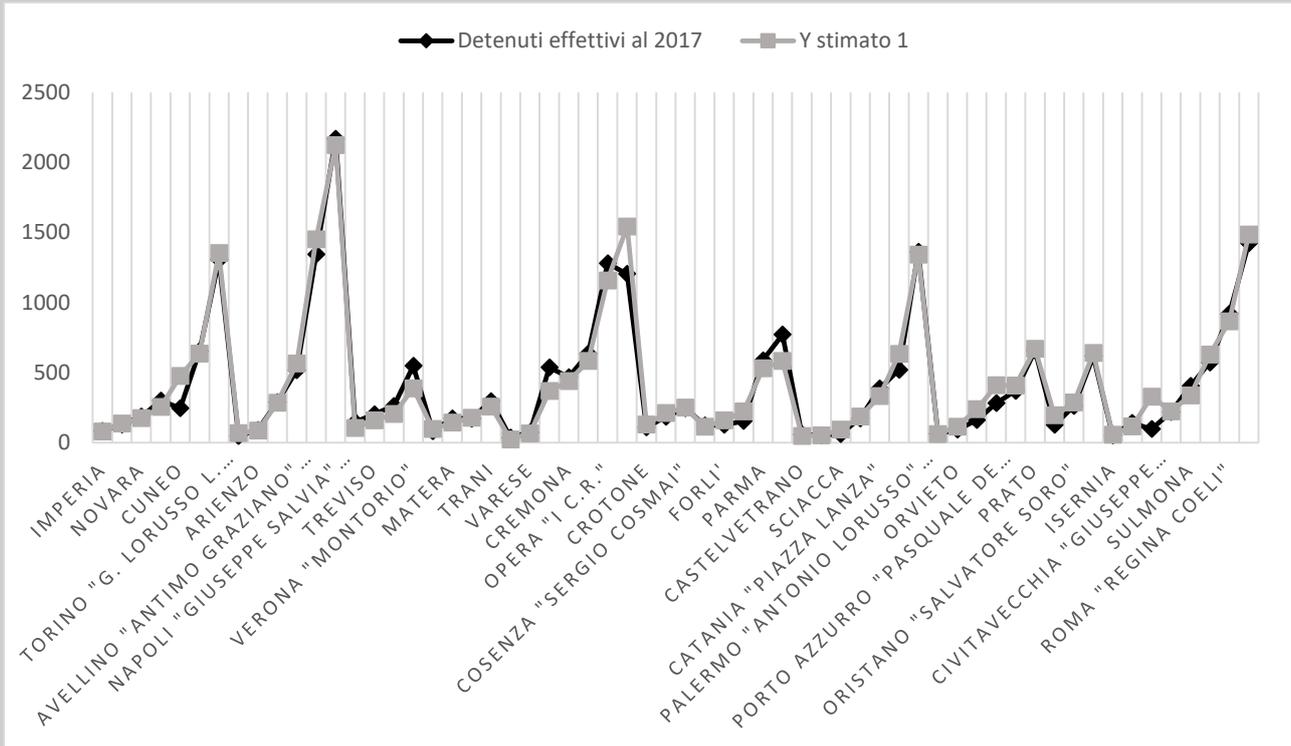
Come è possibile notare, questa volta il modello è stato incrementato con l'utilizzo di una nuova variabile in formato dummy, dunque il software ha restituito 4 diverse funzioni di regressione, ciascuna associata ad una specifica area macro-regionale. Risulta interessante che soltanto per la modalità "Nord-Est" il coefficiente stimato è significativo, mentre per le altre è assimilabile a 0 dal momento che il p-value è >0.05. Dunque, per ogni variazione unitaria dei detenuti presenti (variabile di risposta), le capienze delle carceri subiscono un aumento di 1.13, la criminalità relativa ai reati del gruppo 5 accusa un incremento di 27.5 e, nel caso del Nord-Est, l'intercetta del modello assume un valore aumentato di 126.38 rispetto a quello assunto negli altri 3 casi. Anche per questo modello<sup>10</sup> i risultati sono soddisfacenti dal momento che l'R<sup>2</sup> adjusted, oltre che il p-value generale, assumono valori assimilabili a quelli dei modelli precedenti.

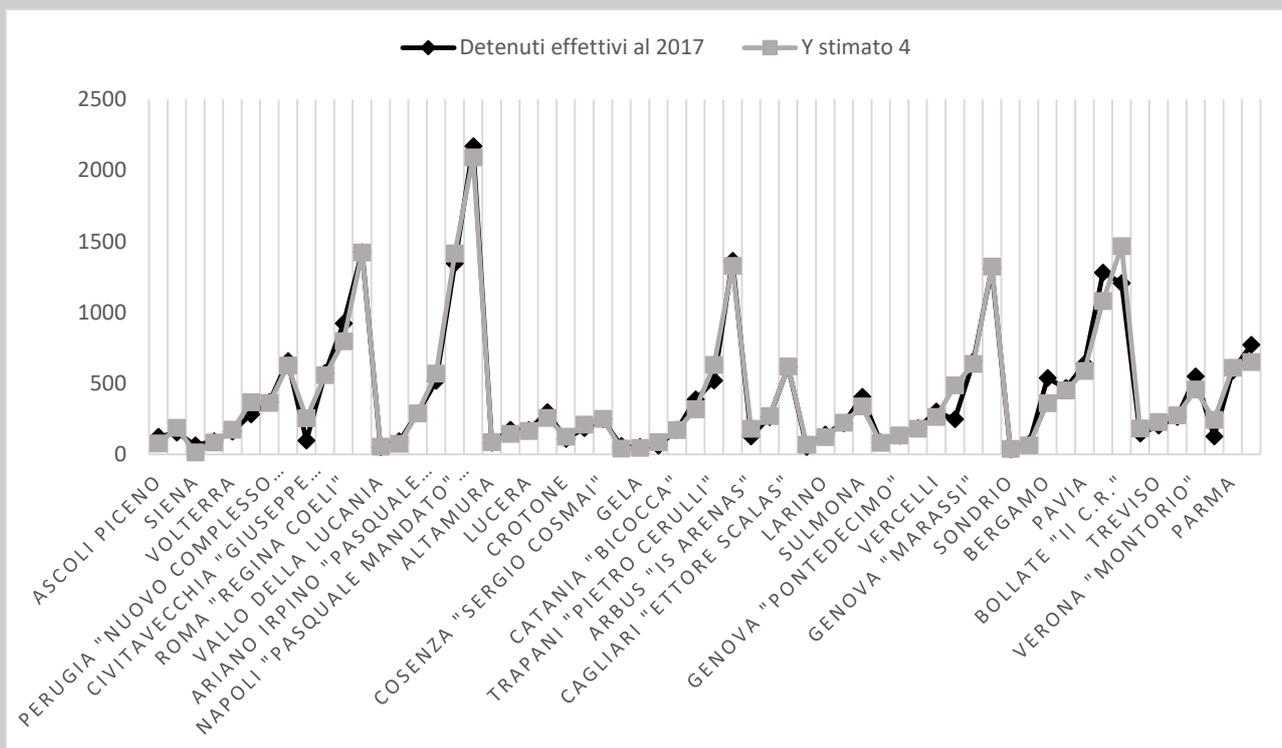
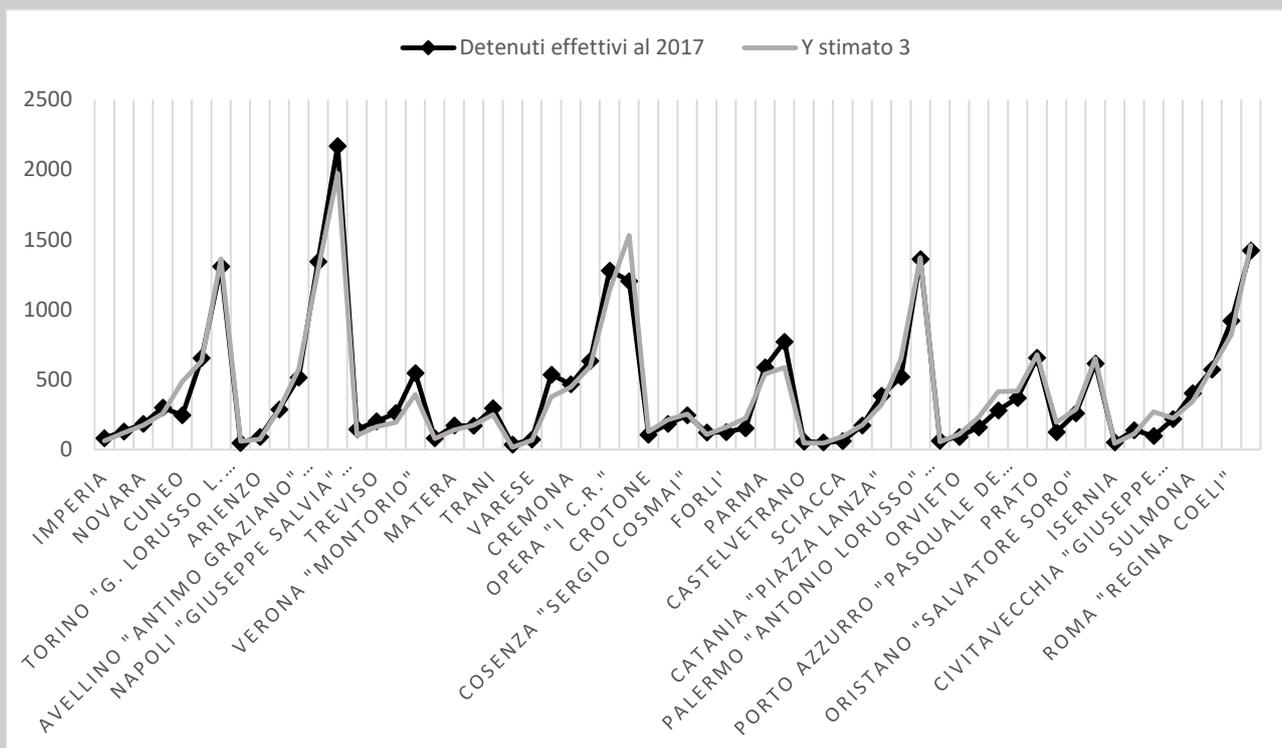
---

<sup>10</sup> È possibile prendere visione del dataset generale in cui sono contenute tutte le modalità delle variabili analizzate associate alle unità statistiche campionate visionando l'allegato D. Si può invece prendere visione dei modelli di regressione costruiti e dei valori stimati, nonché degli scostamenti rispetto ai valori osservati, visionando l'allegato D.

## I risultati

Al fine di verificare l'attendibilità e l'aderenza alla realtà dei modelli sviluppati, sono stati confrontati i risultati stimati dalle diverse funzioni di regressione con i valori osservati relativamente alle unità statistiche campionate. Di seguito la rappresentazione grafica dei confronti operati.





Alla luce di quanto sopra, il modello preferibile risulta essere il quarto dal momento che lo scostamento medio è del 4.2% (compresi gli outliers) e mediano inferiore all'1%. Con i dati disponibili il modello riesce a prevedere, con un errore individuale inferiore al 30%, 51 casi su 60 (85%). La funzione di regressione dunque si adatta quasi perfettamente alla realtà e risulta

attendibile per scopi predittivi, sempre considerato il fatto che non sopraggiungano modificazioni sostanziali al corrente sistema giuridico penale sia in termini di fatti previsti dalla legge come reato che in termini di scarcerazioni preventive a seguito di provvedimenti parlamentari (es. indulti), il che renderebbe necessario rivedere il modello per degli aggiustamenti del caso. Nonostante la bontà della funzione di regressione, è necessario specificare che i dati raccolti ed utilizzati sono migliorabili: sarebbe preferibile che questi ultimi provenissero da un'unica fonte con la più alta tempestività possibile e raccolti direttamente con dettaglio provinciale al fine di avere risultati ancora più precisi.

Detto ciò, dev'essere chiaro che questo impianto di regressione non intende descrivere quali variabili spiegano in modo particolare il sovraffollamento delle carceri, bensì l'affollamento in termini assoluti. Chiaramente, per ogni singolo istituto, confrontando il risultato ottenuto dal modello di regressione con il dato sulla capienza è sempre possibile determinare se quel carcere si trova - o si troverebbe nell'anno di riferimento - in una condizione di sott'affollamento o sovraffollamento.

Un possibile sviluppo di questo strumento sarebbe l'impostazione dello stesso in modalità logistica dove la variabile di risposta sarebbe rappresentata da "Tasso di sovraffollamento" con modalità 0 e 1, dove 0 indicherebbe un'alta probabilità di avere sotto-affollamento o raggiungere la capienza massima e 1 un'alta probabilità di avere sovraffollamento.

Un ultimo appunto riguarda la variabile "capienza delle carceri". È emerso, durante l'analisi delle correlazioni tra tutte le variabili considerate, che la suddetta presenta un nesso, a volte più forte ed altre meno, con: la popolazione residente nella provincia dove insiste il carcere in esame, la criminalità, l'indice d'integrazione 1, i residenti stranieri, gli occupati stranieri, gli stranieri iscritti a scuola, l'indice delle condizioni culturali 1 e 2, la popolazione con meno di 5 anni d'istruzione, la popolazione con almeno 5 anni d'istruzione. Alla luce di queste riflessioni, è possibile che - ed è stato ritenuto così in questo lavoro - la variabile "capienza delle carceri" sia una sorta di aggregato latente portatore delle stesse informazioni apportate singolarmente dalle altre variabili. Probabilmente dietro il nesso illustrato si

nasconde una parte del ragionamento di coloro che progettano gli istituti penitenziari. Non essendo disponibili dati più precisi in merito questo è soltanto un assunto ipotetico, meritevole però di un approfondimento specifico soprattutto in forza di un fatto curioso: la “criminalità” seppur correlata alla “capienza delle carceri” non ha costituito un problema di multicollinearità nello sviluppo dei modelli di regressione illustrati, dunque verrebbe da pensare che tra quelle menzionate sia l’unica con una correlazione avente un p-value  $<0.05$ .

## Riferimenti

### Allegati

Allegato A: variabili e relative informazioni pratiche. Cliccando due volte sull'icona si potrà visualizzare una tabella con l'elenco completo delle variabili analizzate e relative informazioni pratiche (fonti, periodo di riferimento, area di riferimento).



Allegato B: “piano di campionamento.xlsx”. Cliccando sull'icona si potrà visualizzare il file .xlsx incorporato nella presente relazione e contenente: lista di campionamento, schema di campionamento, campione estratto.



Allegato C: analisi descrittive delle variabili passate alla lente d'ingrandimento. I dataset relativi alle analisi qui proposte non sono allegati dal momento che tutti i dati necessari all'indagine generale sono presenti al successivo allegato.



Condizione  
Sociale.pdf



Criminalità.pdf



Integrazione.pdf



Procedimenti  
penali.pdf



Condizioni  
Economiche.pdf



Indice del clima.pdf

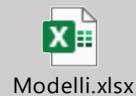


Persone.pdf

Allegato D: “dataset generale.xlsx” contenente tutte le informazioni acquisite in riferimento ad ogni unità statistica campionata. Cliccando sull'icona sarà possibile visualizzare il file incorporato nella presente relazione.



Allegato E: “modelli.xlsx” contenente i modelli di regressione sviluppati ed i relativi risultati associati a ciascuna unità statistica campionata. Cliccando sull'icona sarà possibile visualizzare il file incorporato nella presente relazione.



### *Bibliografia*

Di Cecca, D. (2018). XV rapporto sulle condizioni di detenzione. Tratto da Antigone: <http://www.antigone.it/quindicesimo-rapporto-sulle-condizioni-di-detenzione/il-lavoro/>.

Materia, S. (2016). Torna il carcere, XIII rapporto. Tratto da Antigone: <https://www.antigone.it/tredicesimo-rapporto-sulle-condizioni-di-detenzione/04-diritto-al-lavoro/>.

Ministero della Giustizia. (2019). Statistiche. Tratto da Ministero della Giustizia. Percorsi chiari e precisi: un tuo diritto.: [https://www.giustizia.it/giustizia/it/mg\\_1\\_14.page](https://www.giustizia.it/giustizia/it/mg_1_14.page)

### *Sitografia*

[www.antigone.it](http://www.antigone.it) (associazione Antigone)

[www.giustizia.it](http://www.giustizia.it) (Ministero della Giustizia)

[www.webstat.giustizia.it](http://www.webstat.giustizia.it) (Data-warehouse Ministero della Giustizia)

[www.istat.it](http://www.istat.it) (Istituto Statistico Nazionale)

[www.dati.istat.it](http://www.dati.istat.it) (Data-warehouse ISTAT)

[www.ilsole24ore.com](http://www.ilsole24ore.com) (quotidiano Il Sole 24 Ore)

[lab.24.ilsole24ore.com](http://lab.24.ilsole24ore.com) (Data-warehouse de Il Sole 24 Ore)

[www.bancaditalia.it](http://www.bancaditalia.it) (Banca d'Italia)

[www.insulaR.it](http://www.insulaR.it) (Guida all'utilizzo di software per l'analisi statistica)

[www.insidecarceri.com](http://www.insidecarceri.com) (Reportage fotografici di Antigone)

[www.ilfattoquotidiano.it](http://www.ilfattoquotidiano.it) (testata giornalistica particolarmente attiva su tematiche penitenziarie)

### *Tesi consultate*

Pannozzo A. (2019) *Il governo delle carceri: funzioni ed organizzazione dell'Amministrazione penitenziaria*. Università degli Studi di Roma Tre, Scienze per il governo e l'amministrazione. (Acquistabile al link: <https://www.brocardi.it/tesi-di-laurea/governo-delle-carceri-organizzazione-funzioni-dell-amministrazione/143.html>)

### *Fonti normative*

L. 354/1975 (Ordinamento penitenziario O.P.)

D.P.R. 230/2000 (Regolamento d'esecuzione dell'O.P.)