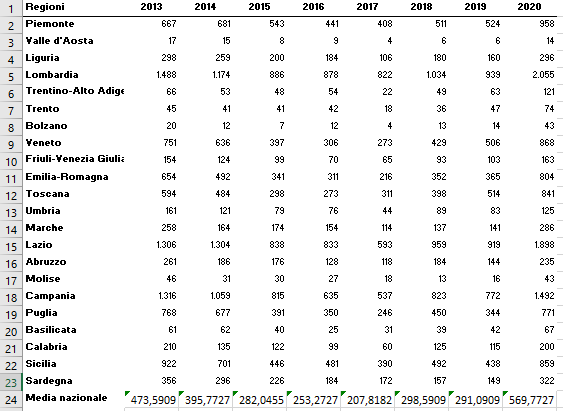
# Statistica descrittiva univariata

Per l’analisi del fenomeno in esame si considerano i dati relativi agli utenti del numero antiviolenza 1522 effettuate nei mesi di marzo-giugno suddivisi per regione ed anno (2013-2020). In particolare, verranno esaminate nei dettagli le curve relativi ai dati della regione Campania e la media delle chiamate effettuate sull’intero territorio nazionale.



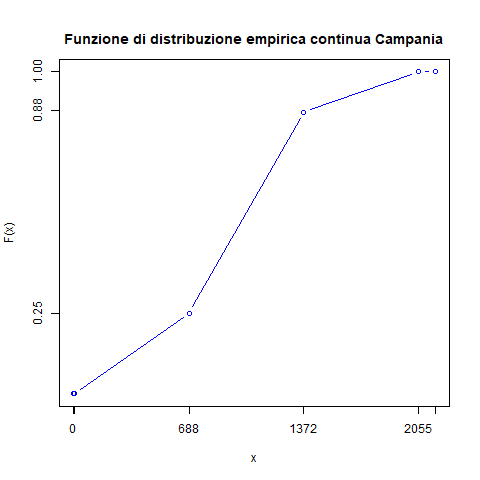
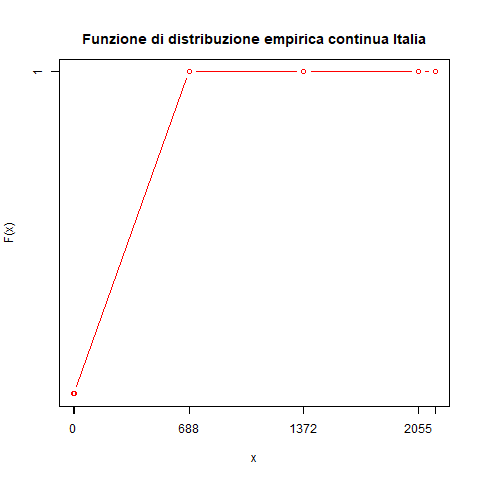
Sono stati creati due barplot, uno relativo ai dati della media nazionale e l’altro relativo ai dati in Campania.





## Funzione di distribuzione continua

Per calcolare la funzione di distribuzione continua sono stati suddivisi le osservazioni sono state suddivise in 3 classi: C1 = [4, 688), C2 = [688, 1372), C3 = [1372, 2055]. Sono stati quindi creati i grafici che mostrano le frequenze di distribuzione continue della Campania e dell’intera nazione. Nel grafico relativo all’Italia si può vedere come tutti i valori siano concentrati nella prima classe.



## Indici di sintesi

Alcuni indici di sintesi utili a descrivere i dati sono media, mediana, moda, varianza, deviazione standard e coefficiente di variazione. Le prime tre sono misure di centralità dei dati mentre le altre misurano la loro dispersione. Nel grafico seguente vengono mostrate le due curve relative ai dati che si stanno analizzando.



Entrambe le curve mostrano una distribuzione di frequenze non simmetrica, in particolare inizialmente sono decrescenti ed hanno un picco massimo nell’ultimo anno 2020. Le due curve sono tra loro abbastanza simili.

Il grafico seguente mostra, invece, i boxplot di entrambi i dati per illustrare alcune caratteristiche della distribuzione di frequenza come centralità, dispersione, forma e la presenza di eventuali valori anomali.



Entrambi i boxplot rivelano la presenza di asimmetria nei dati in quanto le distanze tra primo e terzo quartile dalla linea della mediana sono molto diverse tra loro.

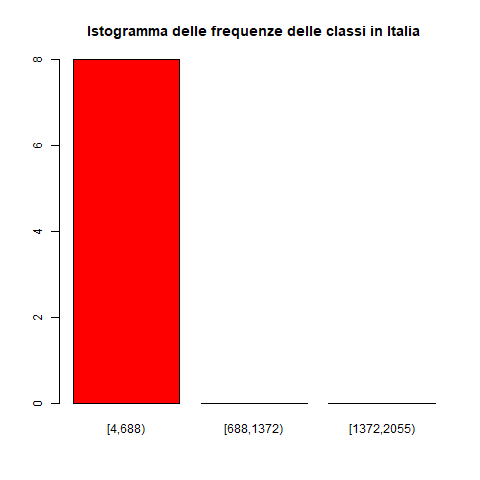
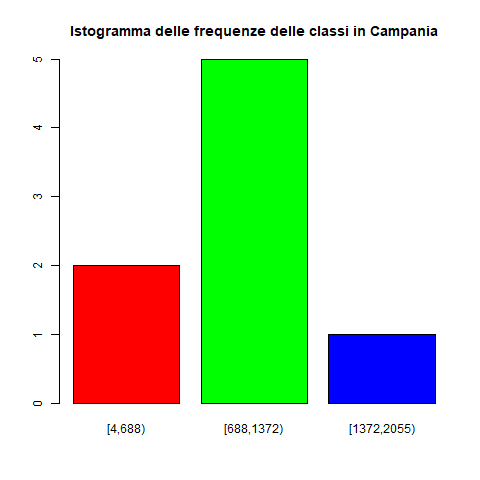
Utilizzando la funzione summary in R è possibile calcolare minimo, massimo, media, mediana, primo e terzo quartile. Dai grafici a barre in alto è possibile inoltre determinare la moda, che in entrambi i casi è associata all’anno 2020. In generale, la curva dei due dati è sostanzialmente simile anche se i dati relativi all’intera nazione sono più bassi in quanto sono calcolati dalla media di tutte le nazioni, che viene fortemente influenzata dai valori bassi presenti in molte regioni con meno abitanti rispetto alla Campania.





Per individuare la classe modale si osservano i barplot relativi alle frequenze dei valori nelle varie classi.

La classe modale per l’Italia è la prima, in particolare tutti i valori sono concentrati in quella classe. Per la Campania invece la classe modale è la seconda.



## Forma della distribuzione di frequenze

**Skewness campionaria**. Entrambe le distribuzioni di frequenze hanno un’asimmetria positiva, la distribuzione di frequenza ha quindi una coda più allungata a destra.



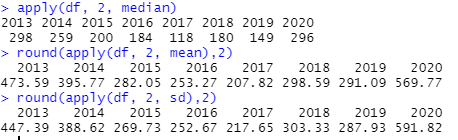
**Curtosi campionaria.** Il valore di entrambe le curtosi campionarie è negativo quindi entrambe le distribuzioni di frequenza sono meno piccate di una distribuzione di frequenze normale standard.



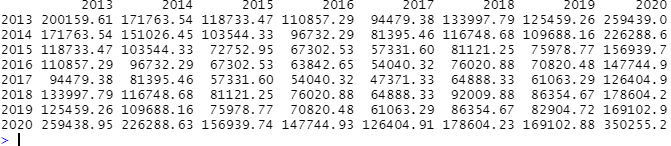
# Statistica descrittiva bivariata

È stato utilizzato il modello di regressione lineare multipla per spiegare la relazione le variabili indipendenti: 2013, 2014, 2015, 2016, 2017, 2018, 2019 e la variabile dipendente: 2020

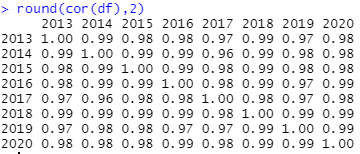
Valore degli indici di posizione e di dispersione relativi alle variabili:



Successivamente, è stata calcolata la matrice delle covarianze



La matrice delle correlazioni che contiene tutte le correlazioni lineari tra le coppie di variabili.

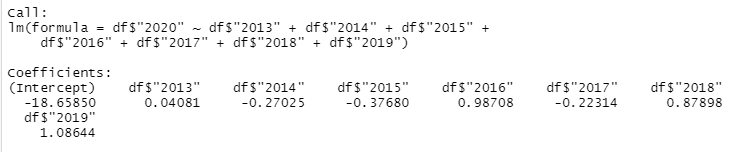


Si nota che esiste una forte correlazione lineare tra tutte le variabili considerate.

Il seguente grafico visualizza in un’unica finestra tutti gli scatterplot ottenuti mettendo in relazione le varie coppie di variabili.



Utilizzando il modello di regressione lineare multipla si ottiene:



Da cui si ricava che l’intercetta è -18.65850 e i regressori sono: 0.04081, -0.27025, -0.37680, 0.98708, -0.22314, 0.87898, 1.08644.

Pertanto il modello di regressione stimato è:

y=-18.65850+ 0.04081x1 + …

I segni dei regressori b1, b4, b6, b7 sono positivi: questo indica che all’aumentare del numero di utenti nel 2013, 2016, 2018 e 2019 aumenta il numero di utenti nel 2020. Mentre i regressori b2, b3, b5 sono negativi quindi all’aumentare del numero di utenti nel 2014, 2015, 2017 diminuisce il numero di utenti nel 2020.

Il regressore di 2013 è prossimo allo zero, questo indica che il numero di utenti nel 2013 non incide in maniera significativa il numero di utenti nel 2020.