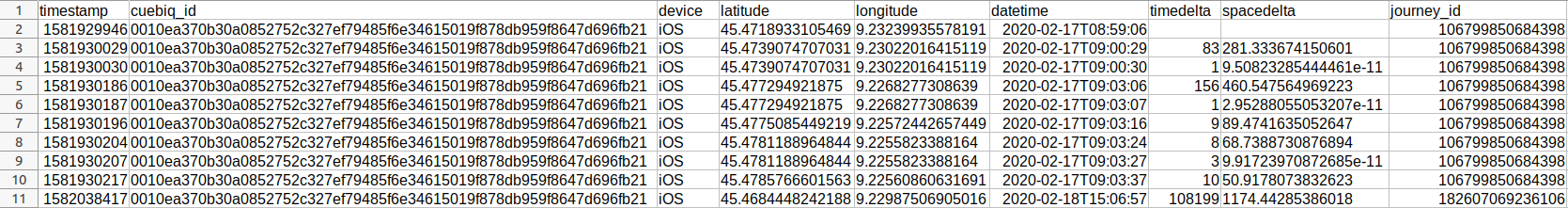
COVID AND MOBILITY

* general problem
* the data
* reasearch o jectives and expected result
* analysis tools used and result obtained

How did the pandemic affect the mobility in Lombardia?

In order to answer this question we are analizing a dataset colletting information about people moving in this region.

Our dataset contains GPS locations collected from February 17th 2020 to May 5th 2020 from around 50k users of some app providing their data to Cuebiq.



As you can see, each observation is a record of the location of a user: in particular it is characterized by 10 variables.

The first one is cuebic­\_id, a univoque code identifying each user (i.e. each mobile phone and its operating system)

(we have modified the one shown for privacy reasons)

We than have latitude, longitude for every single observation.

Datetime and timestamp are two different ways to measure time.

Spacedelta measures the distance (in meters) which occurs between one observation and the previous.

Finally we have journey\_id. Indeed every sequence of observations constitues the so called journey if the time interval between two subsequent observations is less than one hour.

(slide esempio iene esselunga)

EXAMPLE

Our main questions are:

1. There has been a change in people’s habits of movements between the end of february and the beginning of lockdown?
2. Had the government’s dpcm (government restrictions) an influence on these?
3. Is it possible to make a prediction of pandemic trend based on movements? Is there a correlation between the number of daily infected and the number of people going around in the previous days?

**According to the previous questions**, can we conclude whether the lockdown is a successfull rescriction or not?

(histogramma viaggi per giorno con attenzione a un’variazione brusca tra fine febbraio e inizio marzo ? (Lockdown improvviso)

Variazione meno brusca tra fine aprile e inizio maggio ? (Riapertura graduale)?

* Is it possible to make a prediction between movements and pandemic trend (infected and victims)?
* Luoghi frequentati (DA VEDERE)/// heatmap???????

hp gaussianità: non gaussiani

1. trasformazione box-cox lambda= 7.840166 ---> dati gaussiani
2. weekend meno spostamenti.  dividendo dataset in giorni settimanali e weekend vengono entrambi gaussiani.

Anova:

Divido in g gruppi, dove ciascun gruppo è due settimane di osservazione.

p=1 (univariato)-> viaggi/giorno (->> 21 osservazioni per gruppo  (ultimo gruppo ha 14 osservazioni in attesa dei dati di maggio/giugno))

abbiamo evidenza di rifiutare H0: campioni provenienti da gruppi diversi hanno diversa media

Da intervalli di confidenza di evince che le medie tra gruppo 3 e 4 non differiscono in modo significativo, le altre sì

quindi faccio simBonf per vedere tau(i)-tau(j) cioè come è la variazione tra due gruppi (i.e. tra due settimane e altre due settimane)

Per tutte le persone, capire se è preso lo stesso orizzonte temporale, e in caso contrario cercare di fare una distribuzione della quantità di orizzonte temporale analizzato.

Ci chiediamo, se non abbiamo dati possiamo concludere che l’individuo è effettivamente fermo?

* confrontare 24 febbraio fino a 7/9 marzo (gente in vacanza)
* 4 maggio riapertura

Ricerca su internet dpcm

Trovare grafico curva contagi

Per ogni persona abbiamo lo stesso periodo di osservazione ? Se non abbiamo osservazioni vuol dire che sono fermi ? No, mi sa che è perchè non sanno usato app per cui cuebiq rileva la loro posizione.

Cfr contagiati/morti covid e spostamenti (-> regressione lineare)

Punto di partenza coincide con punto di arrivo ->> passeggiata del cane

Altrimenti ->> lavoro

-  heatmap per cubic.id per capire in quali zone prendiamo le persone (observations) maggiormente

Variazione brusca tra fine febbraio e inizio marzo ? (Lockdown improvviso)

Variazione meno brusca tra fine aprile e inizio maggio ? (Riapertura graduale)

Dataset sul numero di spostamenti giornalieri:

hp gaussianità: non gaussiani

1. trasformazione box-cox lambda= 7.840166 ---> dati gaussiani
2. weekend meno spostamenti.  dividendo dataset in giorni settimanali e weekend vengono entrambi gaussiani.

Anova:

Divido in g gruppi, dove ciascun gruppo è due settimane di osservazione.

p=1 (univariato)-> viaggi/giorno (->> 21 osservazioni per gruppo  (ultimo gruppo ha 14 osservazioni in attesa dei dati di maggio/giugno))

abbiamo evidenza di rifiutare H0: campioni provenienti da gruppi diversi hanno diversa media

Da intervalli di confidenza di evince che le medie tra gruppo 3 e 4 non differiscono in modo significativo, le altre sì

quindi faccio simBonf per vedere tau(i)-tau(j) cioè come è la variazione tra due gruppi (i.e. tra due settimane e altre due settimane)