# Footfall Index

**Data**

For any given month M, we got the one-hour adjusted counts directly from LDC servers.

**Methodology**

The proposed index, measures the relative change in footfall from one month to another as follows:

where b = Total footfall at month , a = Total footfall at base month , .

The challenge is the actual construction of b and a, because

1. The number of sensors is not the same between any two given months (at this stage of the project there are always more sensors in bthan in a);
2. A single sensor could be measuring *H* hours in month and *K* hours in month , with *KH*;
3. In a particular month, some sensors can be considered just as white noise, because they may have only a few valid measures for a variety of reasons and,
4. Really high counts due particular locations, for example, sensors located near phone shops.

These discrepancies make, in principle, b and a incomparable between each other.

To solve this, we proceed as follows:

1. Define , where are the total number of valid hours for sensor d in respectively, and are the counts at sensor *d* at hour *i.* Basically, is the sum of all the footfall in a single month at sensor *d*. An hour is valid if it has an actual value recorded. For example, there are sensors with only 16 valid hours in a single day.
2. Calculate the theoretical probability distribution of and :
   1. Discard all sensors skewed to the left of the bulk of the distribution, i.e. those who are 3 times to the left of the standard deviation. In other words, we removed all sensors that didn’t work properly during months b and a, or that worked only for a brief period of time in those months.
   2. For sensors skewed to the right, i.e., sensors which counts are 3 times to the right of the standard deviation, we first verified if their behaviour is the same in past months or if it was just an anomaly. If is the latter, we removed them from the counts, otherwise, we manually verified their locations to see if they are near a noise source (a phone shop for example), and if that’s the case, they are removed.
3. With the remaining sensors, we define b and a as follows:

,

Where came from 1) and are the total number of valid hours in b and a for all the remaining sensors after 2).