**CRIMES, PUNISHMENT AND CONVICTIONS**

**Brief explanatory description of the datasets**

This document briefly describes the data contained in the files “ *authors\_sex\_age\_conditional\_prob*”, “*punishment\_type*” and “*conviction\_length*”. Each file will contain a “readme” sheet where more detailed information on the structure of the given dataset will be provided.

**Estimation of the real number of crimes and baseline probability** **(“*authors\_sex\_age\_conditional\_prob*”)**

This file aims at resolving or mitigating the recurring problem of the use of reported crimes alone as the only source for quantifying the volume of crimes in a given area and to provide a baseline probability of crime commission based on sex and age of the agents. Regarding the first point, many crime categories suffer from the so-called “dark number” problem. The dark number maps the proportion of offences that are not reported to law enforcement. This number varies based on the crime type and the geographical context. Official estimates of the dark number for several crime categories can be retrieved from ISTAT victimization surveys. The most recent accessible data are related to years 2008 and 2009. Our estimation is based on the following steps:

* Download of the ISTAT number of reported offences per crime category for the years 2012-2016 in the province of Palermo
* Comparison of the overall number of categories with the categories for which a dark number figure exists.
* Derivation of the dark number of categories for which we can assume that the dark number is 0 (e.g.: murders, meaning that we assume that all murders are actually reported) and application of the ISTAT dark number to crime categories that are very similar to ones for which the official figure exist
* For all the categories for which the dark number is not available, cannot be derived nor assumed, we have imputed a value computed via a weighted average of dark number and number of crimes for each category, using Palermo province data for the year 2016 as reference. Specifically, the imputed dark number for the remaining categories has been calculated as follows:

where represents the total number of crimes reported for category *k* and represents the dark number for the same crime category. Applying this method, we have obtained and imputed a weighted average value of 0.68.

* Correction of the absolute reported number of offences for each year, via the simple following transformation and computation of the rate calculated per 10,000 inhabitants using the Palermo population for the year of reference retrieved by ISTAT.
* Synthesis of the results in order to calculate the average corrected absolute values () and rates for the 2012-2016 period. We have provided both the figures per crime category and an aggregate number regardless of crime type.

After the estimation of the real number of crimes, we have also defined a baseline probability of crime commission based on the gender and the age class of the individuals. This calculation was made using data referred to Sicily. The sheet “conditional probability” indeed computes these probabilities using data retrieved from ISTAT on the absolute number of Sicilian residents divided by sex and age class. The available classes are the following:

* Up to 13 years
* From 14 to 17 years
* From 18 to 24 years
* From 25 to 34 years
* From 35 to 44 years
* From 45 to 54 years
* From 55 to 64 years
* Over 65 years

For gender *i,* age class *j* andyear *k,* the probability was computed as the ratio between the overall corrected number of crimes and the total number of inhabitants in Sicily. Then, a final figure was obtained using the arithmetic average for all the considered years.

**Description of the Types of Punishment (“punishment\_type”)**

The aim of the file is to provide information on the prevalence of the different types of punishment over the total number of individuals for which a definitive judgment has been issued in the year of reference. Unfortunately, data were only available at national (Italian) level. Data are retrieved from ISTAT and are related to the 2012-2016 period. This dataset also distinguishes between males and females in order to understand the different percentages and distributions of punishments across genders. For each year, we have provided data on the absolute number of females and males sentenced to either a. *house arrest*, b. *community work*, c. (*prison) conviction*, d. *fine*. For each gender we have then calculated the percentage of a specific punishment over the total number of punishment and the rate per 10,000 inhabitants (using ISTAT population data by sex for the year of reference). Furthermore, we have also computed the percentage of males and females out of the total number of individuals sentenced with a specific punishment, in order to understand - for instance - the gender distribution in prisons. Finally, in the “Average” sheet we have calculated the average values of each aforementioned information over the 2012-2016 period.

**Description of the Conviction Lengths (“conviction\_length”)**

The aim of this third file is to give information regarding the distribution of conviction sentences in terms of length. This file relies upon the same ISTAT dataset used for the “punishment\_type” file, filtering out all sentenced that were not convictions and focusing specifically on the different lengths of the sentence. There were nine available categories (that might be merged and reduced if preferred), namely:

1. Conviction in prison up to a month;

2. From one month and one day to three months;

3. From three months and one day to six months;

4. From six months and one day to one year;

5. From one year and one day to two years;

6. From two years and one day to three years;

7. From three years and one day to five years;

8. From five years and one day to ten years;

9. More than ten years and one day.

For each of these categories (and again distinguishing by sex), we have calculated the percentage of the given category over the total number of convictions; the rate per 10,000 inhabitants (using ISTAT population data by sex for the year of reference) and, finally, the prevalence of each sex in each conviction length category. Finally, in the “Average” sheet we have reported the average values of all these figures over the 2012-2016 period.