Final Assignment

Pietro Romozzi & Federico Colombo

20 novembre 2014

Abstract:

With the present work, the authors intend to create a descriptive model of economical, social and environmental characteristics of a territory, aiming to assess the impact of these factors on suicide rates. The model will be tested on Italian counties (110 observations).

Introduction:

Topic

Several papers have tried to propose an explanatory framework to describe how climate conditions affect mental health. Climate may affect individuals' mental health both directly through extreme climate change, which exposes people to trauma, or indirectly, as some climate's characteristics may affect physical activity and mental health. Extreme temperature or humidity rates have an influence on individuals' physical health as it becomes more energy consuming to do physical activity, to work or to travel, but also mental health could be negatively affected as climate erodes physical environments, which subsequently damage social environments and thus, mental condition. Health is defined by the World Health Organisation as a 'state of complete physical, mental and social wellbeing and not merely the absence of disease or infirmity' (World health organisation, 1948). Mental health is one of the key components that define general health and it can be seen as an alteration in thinking, mood, behaviour, and associated distress or impaired functioning. Usually, problems related to mental condition can be considered to be either uncommon (schizophrenia, anxiety and personal disorders) or common (depression, dementia, physiological distress). The WHO showed that among almost 90% of people who died by suicide in high-income countries at least one mental disorder was present, while those affected by more than one mental disorder were experiencing significantly higher risks of committing suicide. The relationship between mental health (and the related suicide rates) and climate still needs to be studied into details, but some scholars identified extreme heat as a factor that is often associated with a general increase in aggressive, criminal and suicide rates (Bouchama A. (2007)). Humidity has also been associated with mental functioning, especially with poorer concentration and elevated fatigue. Climate change is a phenomenon which is expected to have severe consequences on the mental health of millions of people too: the threat to physical health or the potential loss of home, jobs and familiars due to catastrophic climate changes negatively affects mental condition. Sunlight is another factor, which has been documented to have direct effects on climate related health problems such as depressive symptoms, anxiety disorders, and other mental problems usually peaking in the cold season when there is limited sunlight (Magnusson (2007)). The debate about the impact of climate change on human health has only very recently included considerations of mental health. Climate can have an influence on our psyche in a varieties of ways: it can directly inflict more and more natural disasters on human settlements, which, as evidence suggests, increases anxiety-related or chronic severe mental problems; climate can also increase the risk of physical injuries or health problems, which are correlated with mental health; finally, climate can endanger the natural and social environment on which people depend for their wellbeing.

Relevance

One obvious reason why this work is potentially interesting is that mental disorders and suicide are nowadays often discussed both among insitution in health's field and media. Moreover, our focus on environmental and climati characteristics grounds on a recent body of studies which recognizes the link between these dimentions and mental disorders. These variables are going to be included in addition to more traditional

variables (i.e. related to economic dimensions) to develop a wider model of analysys of mental disorders. As a consequence, the broad spectrum of this study makes the present work interesting and innovative. Another valuable point is the regional focus which allows to underscore, in the end, contingent specific characteristics of the territory rather than the whole country. Therefore, bringing the analysis at a smaller level, makes easier for policy makers to deal with these issues at a local and so more manageable level, as stated before.

According to the latest WHO report concerning suicide prevention (see the latest report by WHO 2014), "Globally, among young adults 15-29 years of age suicide accounts for 8.5% of all deaths and is ranked as the second leading cause of death (after traffic accidents). Among adults aged 30-49 years it accounts for 4.1% of all deaths and is ranked the fifth leading cause of death". This statement displays clearly the relevance of this issue, which is also greater considering that these estimation are often underrated because of the difficulties in recognizing sucide as a cause of death - will of committing suicide could be not recognized, families could prefer not to speak about it, and so on (see the note by ISTAT 2012). The above mentioned paper, moreover, is considered by several OECD countries as a strong stimulus to move further in suicide prevention. Among these countries, we are going to focus on Italy, particularly at regional level. Indeed, our convintion is that a focus on smaller than countries environments, does help in assessing how factors affecting suicides impact differently within the same country, thus highlighting a possible variance among regions which is overlooked by country-level aggregate data. In addition to this "micro-level" focus, we also intend to stress some environmental and climatic characteristics which could contribute to create "suicide-genic" environment. Recognizing these characteristics will be helpful for setting efficient prevention strategies, and, especially considering the regional focus, it will be more easy for policy makers to design policies tailored to the specific regional reality. Especially the climatic characteristic mentioned, according to recent literature, has turned to be correlated with mental health and, trough it, with suicide. These are often not considered, as some auhors like Berry, Bowen, and Kjellstrom (2009) underline, therefore we believe that including them in the present paper could deliver interesting results and perhaps stimulate further research in this branch.

Research questions:

The paper will focus on the relation between suicidal behaviours (an extreme consequence of mental disorders) and various factors concerning climate (Magnusson 2007). In particular, we want to test if the following hypothesis hold true:

- Suicide rate in a certain Italian region increases when the average temperature is low (Bouchama A. 2007).
- Suicide rate increases when precipitations happen to be frequent.
- Suicide rate increases when GDP per capita is high (OECD 2013).
- Suicide rate increases according to economic inequality (represented by the gini index).

The hypothesis we want to test originate from a careful review of the relevant literature regarding mental disorders and climate change. The paper will focus on the relation between suicidal behaviours (an extreme consequence of mental disorders) and various factors concerning climate.

Methodology:

Sources and data gathering

For the scope of this work, regional data were necessary; in particular we looked for data about weather (i.e. temperature and precipitations), GDP per capita and inequalities (i.e. gini index). Fortunately, most of the data required were founded in the Italian National Institute of Statistics (ISTAT) database, or in other databases of ISTAT-related agencies. Accessing to the ISTAT database was fairly easy especially because

data were clear and well structured. Nevertheless importing data in R, where we intended to carry out our analysis, was more problematic. One necessary step was to export relevant data as .cvs files from the ISTAT website, and then proceed to import them in R using the *read.csv* command. Data from the ISTAT database were not tidy though, as in tidy data:

1. Each variable forms a column.
2. Each observation forms a row.
3. Each type of observational unit forms a table.

Therefore, we have to put our data trough a tidying process, that is structuring our datasets to facilitate analysis (Wickham 2014).

To do so we partly used excel's functions, since the data got from ISTAT were not so many and fairly manageable. Then we concluded our tidying process in R, using more complex functions which Excel does not allow to use easily. What follows is the final part of the tidying process we did on R. As the reader could see, at first we cleaned data for suicides, then we merged them with other data we have cleaned and merged together before, in R as well.

Considered variables, data gathering, software used:

Variable *suicide rate* is defined by the WHO as the standardized number of deaths by suicide every 100.000 people occurred in a certain country. Since the focus of this paper is on intra-regional differences in Italy, data from the Italian National Institute of Statistics (ISTAT) will be used. Data regarding our five specific case studies are provided.

Variable *solar radiation* measures the level of solar irradiance in megajoul (MJ) per square meter. The Italian Air Force offers a metereological service, which is responsible for the measurement of solar irradiance on the Italian territory. They also provide documentation gathered from various metereological units, unfortunately not for every single Italian region. Because of that, if regional measurement appears to be too complex for the scope of this work, we might decide not to use this variable or not to consider some regions.

Variable *urban population* measures of the degree of urbanization of a population. It describes the percentage of the total population living in urban areas, as defined by the country. Also in this case, data provided by Istat will be used.

Regarding variable *GDP per capita*, we will use Eurostat's database, where it is possible to find data regarding GDP and GDP per capita for every single region in the European Union. Database is easy to access and to use.

Variable *average temperature* simply corresponds to the value of the annual average temperature in a certain region, while variable *precipitations* measures the annual quantitiy of rain in millimiters. Data are provided by ISTAT (based on Unità per la climatologia e la meteorologia applicate all'agricoltura).

Finally, variable *Gini Coefficient*, the most common measure of income inequality, varies between 0 and 1 and expresses inequality among values of the frequency distribution of income for our target regions. Eurostat provides this typology of data for all European regions.

Data presented above regard:

- Data about climate obtained from CRA-CMA (Department for meteorology applied to agriculture, ISTAT) regarding annual average temperatures, measured in Celsius degrees, and average annual precipitations, measured in millimiters for all years between 2005 and 2009 (ISTAT 2012).
- Data about GDP measured in Euro per capita, found in the ISTAT database as well.

- Data about the Gini Index downloaded from the Istat website. Again, data for the three macroregions we are interested in were available. As expected, Gini's index is slightly higher in Southern Regions compared to the Centre and the North.
- Data about suicide rate, measured as number of suicides on 100000 inhabitants.

Unfortunately, as will be explained in the next chapters, after we have carried out our analysis, we found a result problematic to interpret. Conviced that a possible explanation, might reside in too aggregated data, we opted for running our linear model on regional data instead of grouped-by-macro-area ones. It required the construction of another dataframe, whose data were found in the ISTAT database too, and that we processed directly in Excel since we decided to gather data just for 2009. Further developments of our study will try to include a bigger number of observation concerning italian regions or even a panel data analysis.

We believed that a linear model was an effective way to analyze our dataframe because it could give a clear idea of the impact of the selected indipendent variables on suicides which we chose as to be the dependent variable. We also opted for a semplifications of the model, obtained by removing particularly unsignificant variables. Further information about our analysis and about the interpretation of our model, will be discussed in the following sections. In the meantime, we provide the code we used in R to create our complete and simplified model.

All the data gathered will be processed using "R" (see R Core Team 2014). We use as well some packages developed for R, namely "ggplot2" (see Wickham and Chang 2014) and "repmis" (see Gandrud 2014).

Case studies:

We take into consideration five areas: North-West, North-East, Centre, South and Islands in Italy. The advantage of adopting this categorization is that instead of twenty different case studies (there are twenty different regions in Italy), among which some could be really similar because they are located in the same macro area, we will work with only five. Also the Italian Statistical Agency and Eurostat use this categorization for simplicity and some of the databases we are going to use already include data divided according to this idea of clustering homogeneous Italian regions together.

Preliminary operations

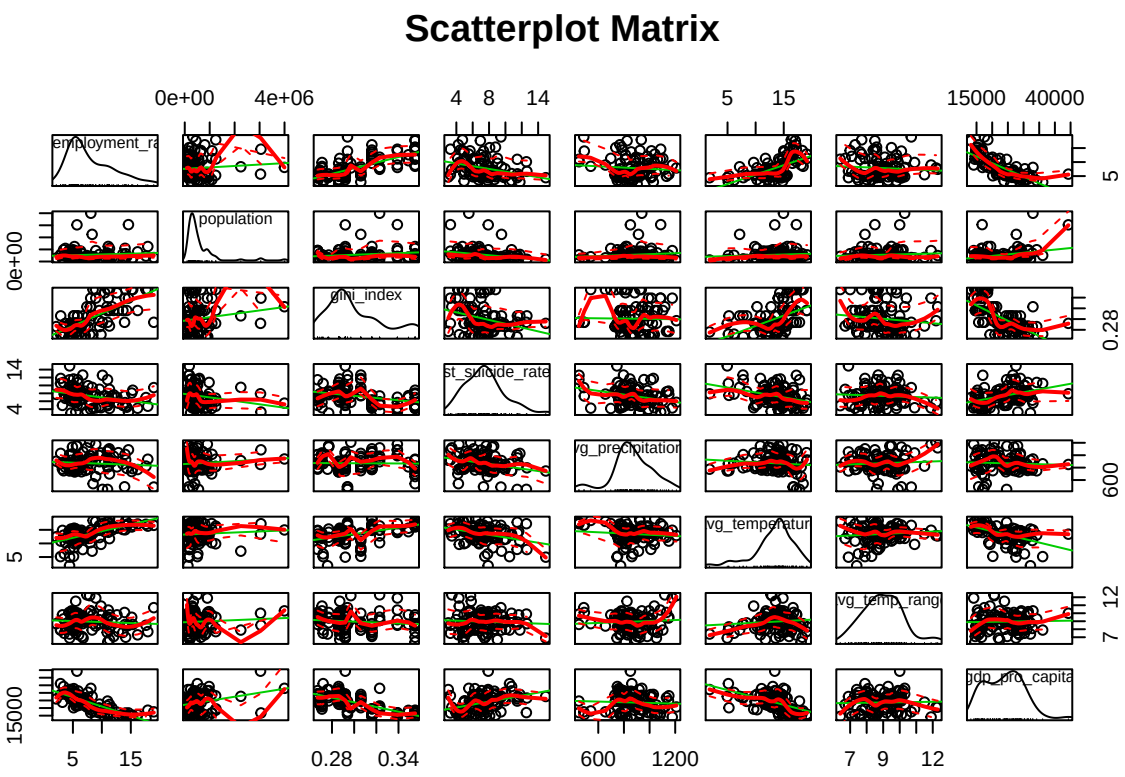
Descriptive Statistics:

```
##
## Please cite as:
##
##  Hlavac, Marek (2014). stargazer: LaTeX code and ASCII text for well-formatted regression and summary
##  R package version 5.1. http://CRAN.R-project.org/package=stargazer


## Loading required package: car
```

Table 1: Data Description

Statistic	N	Mean	St. Dev.	Min	Max
unemployment_rate	107	8.05	3.79	2.13	18.97
population	110	540,306.80	582,856.20	57,329	3,997,465
gini_index	107	0.30	0.03	0.27	0.36
st_suicide_rate	108	7.49	2.68	0.00	19.19
avg_precipitations	103	870.41	145.51	449.00	1,209.30
avg_temperature	103	13.48	3.20	1.80	19.15
avg_temp_range	103	8.95	1.16	6.40	12.30
gdp_pro_capita	110	23,334.55	6,103.13	11,000	44,300



Proposed Model:

The model we want to estimate will therefore be the following:

$$SuicideRate = \alpha + \beta_1 Temp + \beta_2 SolarRadiation + \beta_3 UrbanPop + \beta_4 Precip + \beta_5 Gini + \beta_6 GDPperCap + e$$

A first linear model using all the explanatory variables:

A second, simplified, model:

$$SuicideRate = \alpha + \beta_1 Temp + \beta_2 SolarRadiation + \beta_3 UrbanPop + \beta_4 Precip + \beta_5 Gini + \beta_6 GDPperCap + e$$

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	6.35	5.26	1.21	0.23
gdp_pro_capita	0.00	0.00	3.03	0.00
unemployment_rate	0.13	0.10	1.30	0.20
population	-0.00	0.00	-3.28	0.00
gini_index	4.11	11.44	0.36	0.72
avg_temperature	-0.15	0.08	-1.98	0.05
avg_temp_range	-0.01	0.17	-0.03	0.97
avg_precipitations	-0.00	0.00	-2.67	0.01

Table 2: First Model Estimation

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	7.71	2.85	2.70	0.01
gdp_pro_capita	0.00	0.00	3.20	0.00
unemployment_rate	0.13	0.09	1.41	0.16
population	-0.00	0.00	-3.42	0.00
avg_temperature	-0.15	0.07	-2.02	0.05
avg_precipitations	-0.00	0.00	-2.74	0.01

Table 3: Second Model Estimation

(Intercept)	7.714
gdp_pro_capita	0.000
unemployment_rate	0.132
population	0.000
avg_temperature	-0.149
avg_precipitations	-0.004

Table 4: Table of Coefficients

	2.5 %	97.5 %
(Intercept)	2.049	13.379
gdp_pro_capita	0.000	0.000
unemployment_rate	-0.054	0.317
population	0.000	0.000
avg_temperature	-0.295	-0.002
avg_precipitations	-0.006	-0.001

Table 5: Table of Coefficients

Main findings and conclusions:

References:

Berry, Helen Louise, Kathryn Bowen, and Tord Kjellstrom. 2009. "Climate Change and Mental Health: a Causal Pathways Framework."

Bouchama A., Mohamed G., Dehbi M. 2007. "Prognostic Factors in Heat Wave Related Deaths: a Meta-Analysis." *Arch Intern Med.*, no. 20: 2170–76.

Gandrud, Christopher. 2014. *repmis: A Collection of Miscellaneous Tools for Reproducible Research with R.* http://CRAN.R-project.org/package=repmis