Heatlh concerns in Germany

Vojtěch Sýkora [*1] **Denis Kovačević** [*2] **Nam Nguyen** [*3]

Abstract

Even though modern medicine is able to overcome many obstacles in improving life quality. However, without a healthy lifestyle, even the best health system are of little help. In this paper, we will analyse the most prevalent health concerns in Germany and their development over time. We will do so by using data from multiple sources, including the Global Burden of Disease and the World Development Indicators. Using multivariate linear regression, we are able to identify the most important factors that influence the health of the German population.

1. Introduction

Germany, one of the largest economies in the world with its advanced healthcare system, faces an intriguing paradox: its life expectancy lags behind other high-income countries. This discrepancy, as highlighted in the analysis by Jasilionis et al. (2023) in "*The underwhelming German life expectancy*," poses critical questions about the underlying factors contributing to this phenomenon. Among these, cardiovascular diseases (CVD) emerge as a significant area of concern.

The reasons behind this phenomenon are still undefined. This paper aims to add to the work of Jasilionis et al. (2023) focusing on the most impactful disease out of CVDs and investigating how elements such as an aging population, lifestyle choices, and dietary habits might correlate with the incidence of CVDs.

Firstly, we will provide a brief overview of the most prevalent diseases in Germany in the time period from 1990 to 2019. We use the data obtained from the Global Burden of Disease study (for Health Metrics & , IHME). We will perform a hypothesis test to determine whether the incidence of CVDs is significantly higher in Germany than the global average. After doing so, we introduce the most common CVDs in Germany and compare their incidence to the rest of the world. We also offer insight into the quality of the healthcare system and its success in treating CVDs.

Secondly, we will investigate the correlation between the incidence of CVDs and some of the most common risk factors. In this analysis we used the immense dataset of the World Development Incicators by the World Bank. This dataset provided us with a wide variety of factors to explore while offering data on the specific diseases we were trying to investigate.

Permutation test, correlation testing, multivariate regression

2. Data and Methods

In this section, describe *what you did*. Roughly speaking, explain what data you worked with, how or from where it was collected, it's structure and size. Explain your analysis, and any specific choices you made in it. Depending on the nature of your project, you may focus more or less on certain aspects. If you collected data yourself, explain the collection process in detail. If you downloaded data from the net, show an exploratory analysis that builds intuition for the data, and shows that you know the data well. If you are doing a custom analysis, explain how it works and why it is the right choice. If you are using a standard tool, it may still help to briefly outline it. Cite relevant works. You can use the `\citep` (whole citation in parenthesis) and `\citet` (only year in parenthesis) commands for this purpose (MacKay, 2003).

3. Results

In this section outline your results. At this point, you are just stating the outcome of your analysis. You can highlight important aspects ("we observe a significantly higher value of x over y"), but leave interpretation and opinion to the next section. This section absoultely *has* to include at least two figures.

[*]Equal contribution [1]Matrikelnummer 6636502, vojtech.sykora@student.uni-tuebingen.de, MSc Machine Learning [2]Matrikelnummer 6707752, denis.kovacevic@student.uni-tuebingen.de, MSc Machine Learning [3]Matrikelnummer 6608479, nam.nguyen-the@student.uni-tuebingen.de, MSc Machine Learning.

Project report for the "Data Literacy" course at the University of Tübingen, Winter 2023/24 (Module ML4201). Style template based on the ICML style files 2023.

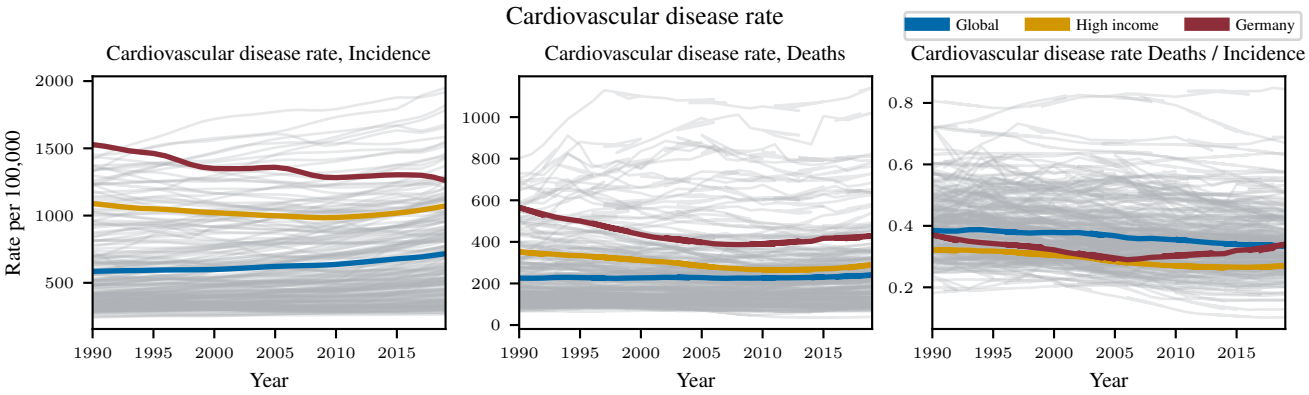

Figure 1. Effect of the cardiovascular diseases on the world over time. From left to right: incidence rate, death rate, and the ratio of death rate to incidence rate. The data is taken from the Global Burden of Disease study (for Health Metrics & , IHME). We specifically focus on the values for Germany in comparison to other high income countries and the world.

4. Discussion & Conclusion

Use this section to briefly summarize the entire text. Highlight limitations and problems, but also make clear statements where they are possible and supported by the analysis.

Contribution Statement

Explain here, in one sentence per person, what each group member contributed. For example, you could write: Max Mustermann collected and prepared data. Gabi Musterfrau and John Doe performed the data analysis. Jane Doe produced visualizations. All authors will jointly wrote the text of the report. Note that you, as a group, a collectively responsible for the report. Your contributions should be roughly equal in amount and difficulty.

Notes

Your entire report has a **hard page limit of 4 pages** excluding references. (I.e. any pages beyond page 4 must only contain references). Appendices are *not* possible. But you can put additional material, like interactive visualizations or videos, on a githunb repo (use links in your pdf to refer to them). Each report has to contain **at least three plots or visualizations**, and **cite at least two references**. More details about how to prepare the report, inclucing how to produce plots, cite correctly, and how to ideally structure your github repo, will be discussed in the lecture, where a rubric for the evaluation will also be provided.

References

for Health Metrics, I. and (IHME), E. Global burden of disease study 2019 (gbd 2019) data resources. http://ghdx.healthdata.org/gbd-2019, 2020. Accessed: [Insert date here].

Jasilionis, D., van Raalte, A. A., Klüsener, S., and Grigoriev, P. The underwhelming german life expectancy. *European Journal of Epidemiology*, 38(8): 839–850, 2023. ISSN 1573-7284. doi: 10.1007/s10654-023-00995-5. URL https://doi.org/10.1007/s10654-023-00995-5.

MacKay, D. J. *Information theory, inference and learning algorithms*. Cambridge university press, 2003.

World Bank. World Development Indicators. https://databank.worldbank.org/source/world-development-indicators/. Accessed: Jan, 2024.