DATA 603 (Fall 2024)

Multi-regression analysis of factors influencing mortality

Question 1 (Name of all group members)

1. Pradeepa Ramamurthy
2. Sai P
3. Gulshan Laskar
4. Yanwei Yu
5. Jennifer Stadtfelder

Question 2 (Introduction): If you have research topics in mind for this project, please briefly describe the topics and any background info I may need to understand the topic. What problems are you going to address? Why are these topics important to you? What are your goals? [2 pts]

The research of influences on adult mortality is an urgent and topical issue that is meaningful for various parties. On the one hand, it is of great importance for health insurance companies and health institutes that are concerned with the optimal development and implementation of prevention strategies and their cost calculation. On the other hand, it is also highly relevant for the pharmaceutical industry, which relies on research in this area for lucrative, innovative and curing therapy and drug development for the prevention and/or treatment of diseases as well as risk factors. This is also underpinned by the research of Hoffman (1998), which demonstrated that mortality can be significantly reduced if awareness is created and access to prevention, medication and clinical treatments is provided based on it.

Accordingly, we will examine the influence on adult mortality based on several potential risk factors outlined in the current literature: year, country, age, pre-existing conditions, population, GDP, developing country, schooling and income composition for resources:

Rogers et al. (2005) concluded that other health conditions and chronic diseases such as HIV, hepatitis or the BMI have an influence on adult mortality, since, for example, the BMI predisposes a person to chronic diseases and long-term complications, or HIV weakens the immune system despite life-prolonging measures. In their work, they also emphasized that one of the main influencing factors alongside unfavorable predispositions is age itself, as the risk of chronic diseases and health problems increases significantly with age, and, thus, also for adult mortality (Rogers et al., 2005). Others argue, however, that although age contributes to a higher risk of adult mortality, it is not the main influencing factor.

Following that, while a study in 1992 has found a positive relationship by age between adult mortality and alcohol consumption (Klatsky et al., 1992), the study by Zhao et al. (1995) suggests that the underlying biological processes in humans indicate that moderate alcohol consumption is beneficial for adult mortality, while excessive consumption is associated with an increased risk of adult mortality (Zhao et al., 1995). However, genetic differences in how different ethnic groups metabolize alcohol imply that these differences are not generalizable and that the combination of country and alcohol consumption should be explored to infer influences on adult

mortality (Cahrtier and Caetano, 2010). Consequently, while there are recent case studies on the impact of alcohol on adult mortality in Europe and the US, we will conduct our study using data from around the world (Machenback et al., 2015; Rostron, 2012).

Furthermore, Pozzer et al. (2023) and Kruk et al. (2018) argue that in particular environmental factors have a significant influence on adult mortality. For example, countries with a high GDP have the opportunity to invest in advanced infrastructure, well-distributed and easily accessible health facilities, environmentally friendly technologies (clean air), clean drinking water and healthy diets, resulting in lower adult mortality than developing countries with a low GDP that lack the means to improve these often-inadequate, poor and critical environmental factors. On the other hand, however, Singer et al. (2001) claims in an extensive literature review that over the last hundred-plus years, research has uncovered a clear link between socioeconomic status (SES) and adult mortality. People with higher SES generally live longer and have lower rates of chronic disease, both in industrialized and less developed countries. As a result, the main SES indicators influencing adult mortality were summarized as income, education and occupational status rather than external factors/which country and whether it is a developing country (Singer et al., 2001). Accordingly, our evaluation and comparison are needed to infer the influences on adult mortality.

Another controversial dependency is population. Countries with a large population often are considered to have a large number of medical facilities and specialists, especially in urban areas, which can lead to better health infrastructure and, therefore, lower adult mortality. In contrast, a large population can also have a negative impact, as high population growth leads to an overload of the healthcare system, which is then no longer able to meet the demand, resulting in longer waiting times for treatment, inadequate medical care and a lack of preventive measures, which in turn increases adult mortality (Rogers et al., 2005). Supporting that, in the case of Covid, it was also found that the infection often spreads faster in countries with high population density, which increases mortality in adults (Wong and Li, 2020). Accordingly, our assessment and comparison are necessary to capture the contrasting influences of GDP and population on adult mortality.

Finally, with the progress of the years, novel, progressive medical research also has a significant impact on average adult mortality. For example, Topalia et al. (2012) found that new immunotherapies and targeted treatments lead to an increased survival rate for certain types of cancer, which increases overall life expectancy (Topalian et al., 2012). Preventive measures such as vaccinations against Covid, HPV and cervical cancer have, thus, been reduced, thereby also reducing mortality in adults. In contrast, progress in years does not necessarily mean progress in medicine and less adult mortality. In contrast, progress in years does not necessarily mean progress in medicine and less adult mortality. For example, covid has significantly increased adult mortality and decreased the willingness to be vaccinated in some population groups during and after covid, which in turn increases the mortality rate in adults (Roy, 2022).

All in all, there are several factors that influence adult mortality. Many researchers disagree on whether certain factors have a significant influence/ or a stronger influence on mortality than others. As a result, we will review and analyze all of these conflicting influences together to get an accurate insight into the influences on adult mortality.

Question 3 (Methodology): Please briefly describe the data you have (or plan to acquire) to help answer the research topics above. Include: what type of variable or variables are included (quantitative, qualitative, etc.), how the variable or variables are measured (the measurement scale), and any other general info you may have on the variable(s) [2 pts]

To answer the above research questions, we access a variety of data in a csv file from the years 2000-2015 from around the world, covering both quantitative and qualitative variables.

Our variable include (Gochiashvili, 2023):

Variable in model	**Variable name**	**Variable type**	**Variable description**
Dependent variable	Life_expectancy	Quantitative	The average life expectancy of people in that country for that year.
Independent variable	Country	Qualitative	The country in which the data was collected.
Independent variable	Regions	Qualitative	There are 9 different regions covering 179 different countries.
Independent variable	Year	Quantitative/Date	In which year was the data collected (from 2000 to 2015)?
Independent variable	Infant_deaths	Quantitative	Number of infant deaths per 1000 inhabitants.
Independent variable	Under_five_deaths	Quantitative	Number of deaths of children under the age of five per 1000 inhabitants.
Independent variable	Adult_mortality	Quantitative	Number of adult deaths per 1000 inhabitants.
Independent variable	Alcohol_consumption	Quantitative	Alcohol consumption is recorded in liters of pure alcohol per capita with 15+ years old.
Independent variable	Hepatitis_B, Measles, Polio & Diphtheria	Quantitative	Represents the percentage coverage of vaccination against the corresponding diseases (Measles etc.) in 1-year-olds.
Independent variable	BMI	Quantitative	Measure of nutritional status in adults, which means average Body Mass Index.
Independent variable	Incidents_HIV	Quantitative	Incidents of HIV per 1000

			population aged 15-49.
Independent variable	GDP_per_capita	Quantitative	GDP per capita (USD).
Independent variable	Population_mln	Quantitative	The total population in that country in millions.
Independent variable	Thinness_ten_nineteen_years & Thinness_five_nine_years	Quantitative	Prevalence of thinness among children or teenagers.
Independent variable	Economy_status_Developed & Economy_status_Developing	Qualitative	Developed or developing country (1 means yes, 0 means no). We will combine these two columns into one variable when we do the regression analysis.
Independen variable	Schooling	Quantitative	Average years that people aged 25+ spent in formal education.
Independent variable	Income composition of resources	Quantitative	The share of each income source or group in the total income of a particular population group or region, expressed as a percentage.

Question 4 (Methodology cont.): Is this your own data set (or the data of someone in the group) or is it "open" or "shared" data? [2 pts]

The dataset we chose for our analysis was part of a project in which various factors were categorized into immunization-related factors, mortality factors, economic factors and social factors affecting life expectancy as part of a multi-linear regression analysis. The information on life expectancy and health determinants for 193 countries was obtained by Deeksha Russell and Duan Wang from the WHO database, while the corresponding economic data was obtained from the United Nations website. As the datasets came from the WHO, there were no obvious errors. Missing/left out data was identified via the “Missmap” function in R and mainly included data on hepatitis B and gross domestic product (GDP) from smaller countries such as Vanuatu, Tonga, Togo and Cabo Verde. The various data files from the open-source databases were then combined into a single dataset and published on the open data source website Kaggle in 2017, indicating that there is no special license to obtain to process and analyze the data further as all the data was based on open sources. We now intend to extend this data analysis/research by focusing on adult mortality and uncategorizing the data to identify more precise determinants rather than using broad categories.

Question 5 (Methodology cont.): Have you distributed the workload among your teammates? If yes, please describe the group members’ workload distribution and responsibilities. [2 pts]

We have distributed the workload among the individual team members. This means that each team member has their own responsibility for certain work but is not limited to that work. As the process progresses, more work can be assigned to everyone, and some initial work can also be removed if it is not needed.

Pradeepa: In the final report, she will outline the steps she took to analyze the logistic regression and determine which model is most appropriate and for what reasons. The justification will be outlined in detail in the final code description.
Sai: He will visualize the data to show and identify correlations and patterns. He will also review the additive and interactive model fit by applying the forward selection procedure and justifying in the final code description which model fits best and why (e.g. adjusted r2 etc.).
Gulshan: She performs data cleaning and checks the fit of the additive and interactive model using the backward elimination procedure. In the final code description, she explains which model is best suited and presents the corresponding arguments (e.g. adjusted r2 etc.).
Jennifer: She finds the best additive regression model using stepwise model selection and explains in the final code description which model fits best and why (e.g. adjusted r2 etc.). She will summarize the results in the interpretation and conclusions part of the final report and will design the presentation.
Yanwei: He determines the most appropriate interaction model by stepwise model selection and argues in the final code description which model is most appropriate and why (e.g. adjusted r2 etc.). He also summarizes the results in the interpretation and conclusion part of the final report and supports the presentation's design.

In the final presentation, the slides and speaking time are divided equally so that everyone is involved.

Reference:

Chartier, K., & Caetano, R. (2010). *Ethnicity and health disparities in alcohol research*. Alcohol research & health: the journal of the National Institute on Alcohol Abuse and Alcoholism. https://pmc.ncbi.nlm.nih.gov/articles/PMC3887493/

Gochiashvili, L. (2023). *Life Expectancy (WHO) Fixed*. Kaggle. https://www.kaggle.com/datasets/lashagoch/life-expectancy-who-updated/data

Hoffman C. 1998. *Uninsured in America: A Chart Book*. The Kaiser Commission on Medicaid and the Uninsured.

Klatsky, A. L., Armstrong, M. A., & Friedman, G. D. (1992). Alcohol and mortality. *Annals of Internal Medicine*, *117*(8), 646–654. https://doi.org/10.7326/0003-4819-117-8-646

Kumar Rajarshi. (2018, February 10). *Life expectancy (WHO)*. Kaggle. https://www.kaggle.com/datasets/kumarajarshi/life-expectancy-who

Kruk, M. E., Gage, A. D., Joseph, N. T., Danaei, G., García-Saisó, S., & Salomon, J. A. (2018). Mortality due to low-quality health systems in the Universal Health Coverage Era: A systematic analysis of amenable deaths in 137 countries. *The Lancet*, *392*(10160), 2203–2212. https://doi.org/10.1016/s0140-6736(18)31668-4

Mackenbach, J. P., Kulhánová, I., Bopp, M., Borrell, C., Deboosere, P., Kovács, K., Looman, C. W., Leinsalu, M., Mäkelä, P., Martikainen, P., Menvielle, G., Rodríguez-Sanz, M., Rychtaříková, J., & de Gelder, R. (2015). Inequalities in alcohol-related mortality in 17 European countries: A retrospective analysis of mortality registers. *PLOS Medicine*, *12*(12). https://doi.org/10.1371/journal.pmed.1001909

Rogers, R. G., Hummer, R. A., & Krueger, P. M. (2005). Adult mortality. *Handbooks of Sociology and Social*

Research, 283–309. https://doi.org/10.1007/0-387-23106-4_11
Rostron, B. (2012). Alcohol consumption and mortality risks in the USA. *Alcohol and Alcoholism*, *47*(3), 334–339. https://doi.org/10.1093/alcalc/agr171
Roy, D. N., Biswas, M., Islam, E., & Azam, Md. S. (2022). Potential factors influencing COVID-19 vaccine acceptance and hesitancy: A systematic review. *PLOS ONE*, *17*(3). https://doi.org/10.1371/journal.pone.0265496
Topalian, S. L., Drake, C. G., & Pardoll, D. M. (2012). Targeting the PD-1/B7-H1(PD-L1) pathway to activate anti-tumor immunity. *Current Opinion in Immunology*, *24*(2), 207–212. https://doi.org/10.1016/j.coi.2011.12.009
Pozzer, A., Anenberg, S. C., Dey, S., Haines, A., Lelieveld, J., & Chowdhury, S. (2023). Mortality attributable to ambient air pollution: A Review of Global Estimates. *GeoHealth*, *7*(1). https://doi.org/10.1029/2022gh000711
Wong, D. W., & Li, Y. (2020). Spreading of COVID-19: Density matters. *PLOS ONE*, *15*(12). https://doi.org/10.1371/journal.pone.0242398
Singer, B., & Ryff, C. D. (2001). *New Horizons in Health: An Integrative Approach*. National Academy Press.
Zhao, J., Stockwell, T., Naimi, T., Churchill, S., Clay, J., & Sherk, A. (2023). Association between daily alcohol intake and risk of all-cause mortality. *JAMA Network Open*, *6*(3). https://doi.org/10.1001/jamanetworkopen.2023.6185

End of Project Checkpoint