How do BMI, GDP and aclcohol influence life expectancy?

Life Expectancy = **-0.21** Alcohol + **0.08** BMI + **0.72** log(GDP) + **1.89** Schooling

1 Background

In an era where public health is increasingly influenced by a combination of individual lifestyle choices and economic factors understanding the determinants of life expectancy has never been more complex and different around the globe. This project seeks to systematically examine how various elements - from personal health behaviours to broader socioeconomic and educational conditions - contribute to the longevity of populations.

2 Methodology

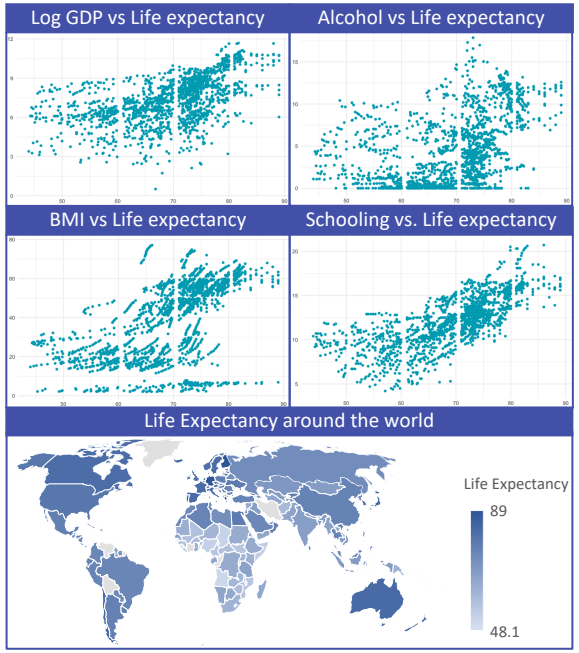
In our project, we focused on analysing life expectancy using data from the World Health Organization. We selected the variables that we believed had a significant impact on life expectancy, such as **Alcohol** consumed per capita, average years of **education** of the population, average **BMI**, and **GDP** in USD. As GDP exhibits very high numbers and showed an exponential relationship when plotting it with Life expectancy, we performed a level-log transformation on it.

Because the dataset contained a times series of several subjects, we performed statistical tests meant for panel data, namely the Robust Hausmann and Breusch-Pagan Test.

We performed the Pooled OLS, Fixed Effects and Random Effects models on the data and compared these models with each other

3 The Dataset

The dataset contains data of almost all countries in the world, starting in 2000 until 2015. It can be observed, that all variables, except for Alcohol have a positive linear relationship with life expectancy.



Africa shows significantly lower life expectancy rates, with Sierra Leone registering the lowest globally at 51 years in 2015. Conversely Europe, North and South America and Oceania, have the highest rates, in which Slovenia takes the lead with 81 years in 2015. The median life expectancy raised from 71 years in 2000 to 73.9 in 2015. However, the average for all countries is approximately 71.62 years, being heavily left skewed.

4 Tests & Estimators

Test	p-value	H_0 - Hypothesis	Conclusion
Robust Hausman	$5.895 * 10^{-8}$	Chose RE model	Chose FE model
Breusch-Pagan (FE)	$< 2.2 * 10^{-16}$	Homoscedasticity	Heteroskedasticity
Breusch-Pagan (RE)	$< 2.2 * 10^{-16}$	Homoscedasticity	Heteroskedasticity
Breusch-Pagan (POLS)	$< 2.2 * 10^{-16}$	Homoscedasticity	Heteroskedasticity

We implemented the Robust Hausman Test as well as the Breusch-Pagan Test for every Estimator. All Tests showed Heteroskedasticity with a significant p-value.

Variable	Pooled OLS	Fixed Effects	Random Effects
Alcohol	-0.21 ***	-0.23 ***	-0.17 ***
BMI	$8.35 * 10^{-2}$ ***	$2.64 * 10^{-3}$	$6.47 * 10^{-0.3}$
log(GDP)	0.72 ***	0.11 *	0.12 **
Schooling	1.89 ***	1.29 ***	1.48 ***
R2-Value	**0.58**	**0.232**	**0.35**

We concluded to use the **Pooled OLS Estimator** as it achieved the highest R2-Value and all Estimators exhibited Heteroskedasticity. With these four variables the model is already able to explain 58% of the data. In this model all variables are highly significant with Schooling being the most important one having an estimator of 1.89, thus having more than double the weight as the second most important variable.

6 Discussion

The huge differences in the R2 value can be explained by the indifference of time associated with the target variable. Because the observed timeframe is relatively small in the context of our dependent variable, we conclude that the dependent variable is independent of time.

It was also really interesting to see that the Covariance of a nation's education and its alcohol consumption was positive correlated to a degree of 6.494 concluding that the more a nation is educated it tends to have a higher alcohol consumption and vice verca. A reason for this can be that people tend to spent their time with more social activities the more wealth they accumulated.

Counterintuvely macroeconomic factors such as GDP and education affect the life expectancy more than individual factors such as Alcohol and BMI.

Ultimately additional research can be conducted if factors not included in the model may be confounding variables. These could be factors such as health care access, quality of nutrition or tobacco and drug use.