An investgation into the use of machine learning techniques as a tool in the study of the factors that affect life expectancy on a global scale

Adriaan Louw (53031377)

June 7, 2019

Abstract

Contents

List of Figures

List of Tables

1 Introduction

Human beings have always had a fascination with longevity. Myths like the fountain of youth or the Holy Grail are a testament to this fact. Today, longevity and causes of mortality are studies by profesionals like Demographers and Actuaries. Trying to determine why some people or group live long lives.

This study investigates the use of Machine Learning techniques in studying the determinants of life expectancy for coutries. Selecting indicators that have shown to have some form of correlation with life expectancy. Their relationship with life expectancy will be investigated using various techniques from Machine Learning and contrasted to various form of regression that is used ubiquitously in the literature. This analysis will seek out to prove the appropriateness of using these machine learning algorithms for use in research to find the exact correlation between these indicators and life expectancy. It is the hypothesis of this study that machine learning techniques like k-Nearest Neighbour and Support Vector Machines will model the indicator/life expectancy relationship better than regression techniques can. Also that these techniques can create more accurate models. In the hope that the causes of long life expectancies in certain countries can be better understood. This study does not aim to prove causation between the indicators chosen and life expectancy, but rather the usefullness of machine learning algorithms as a tool.

Statistical regression techniques are predominantly used algorithm for data analysis. Linear regression (Section 3.2.1) forinstance assumes a linear relationships between the independant variables and the dependant variable. Which might not be the case. This we will discuss in Literature Review (Section 2).

Machine learning is used in medicine Chen & Asch (2017).

life expectancy vs mortality rate?

Machine learning techniques can find relationships in the data that regression analysis cannot Chen & Asch (2017)

Rajkomar et al. (2018) Google uses machine learning to predict in hospital medical events for patients.

Human attempts to mathematically predict life expectance is not a new endeavour. Gompertz (1825) introduced an equation to predict life expectancy, which was modified in Makeham (1860) to create the famous Gompertz−Makeham law.

2 Literature Review

Life expectancy and mortality are 2 related terms. Mortality is generally expressed as a mortality rate. It describes the rate that people die under certain circumstances. Life expectancy (Section 2.2) is the amount of years an individual or group of pople are expected to life. If the amount of people who are dying increses, the mortality rate increases and the life expecancy for people in that group decreases. And visa versa.

Forecasting Mortality in Developed Countries Tabeau 2001

2.1 ?Grossman?

2017 determinants of health: an economic perspective ???? 1972 The Demand for Health: A Theoretical and Empirical Investigation,

Grossman (2000)

2.2 What do we mean by life expectancy?

A life table is a table given for a specific year that contains the probability that a person of a certain age will die in that specific year. Life tables are also called actuarial tables and are used by actuaries in the life insurance industry. Table 1 is an example of a life table taken from Arias (2007). Life expectancy is one element of a life table. Life tables are created by countries and the United Nations

There are 2 types of life tables namely period and cohort life tables. A cohort is a group of people who were born in the same year. A cohort life table will follow a cohort over its lifetime until every member of the cohort has died. A cohort life table requires the mortality information of the cohort over many years. This information is often unavailable. While for a period life table, a hypothetical cohort is created and subjected to current mortality rates. This gives the user of the period timetable a window to see mortality rates at that point in time. This makes period life tables the most common type (Arias et al. 2017).

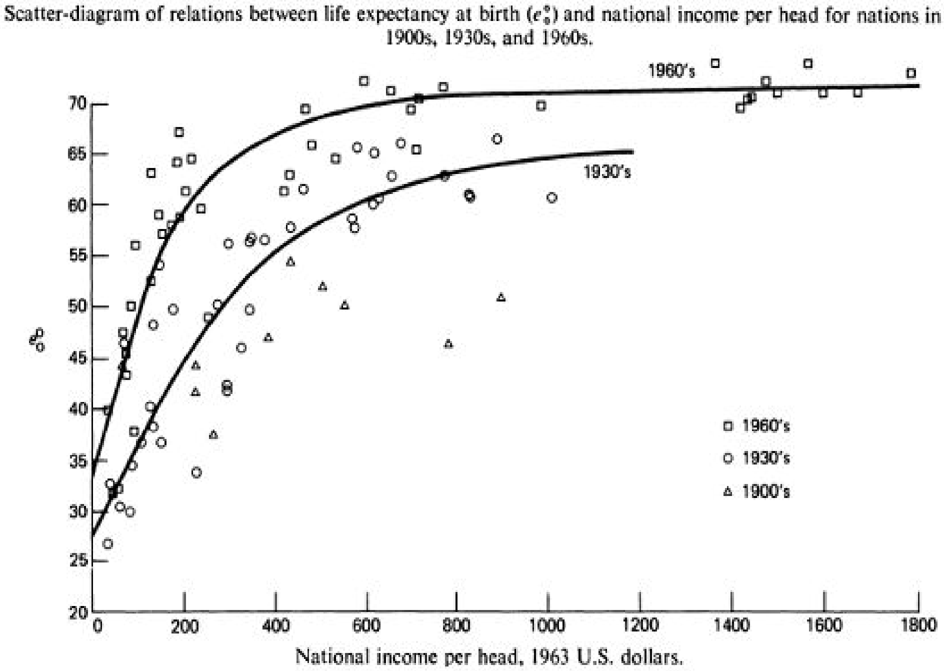

Figure 1: The original Preston curve from Preston (1975)

2.3 Determinants of life expectancy

This section summarises the relationship some of the most well known indicators have with life expectancy.

2.3.1 Income

The relationship beteen income and life expectancy has been given a lot of attention in academic circles (Preston 1975, Hu et al. 2015, Chetty et al. 2016, Oeppen 2019).

Preston (1975) was the first to show the relationship between life expectancy and per capita income. His original curve can be seen in Figure 1. As we can see from Figure 1, for low income countries, life expectancy increases rapidly with per capita income. Whereas in high income countries a small increase in per capita income does not have a large effect on life expectancy.

This relationship has also been shown in more recent studies (Chetty et al. 2016, Oeppen 2019). Even though Shkolnikov et al. (2019) found that in Russia the Preston curve is not an accurate predictor of life expectancy. They found that the actual life expectancy should be "substantially higher" when comparing to the Preston curve predicted value.

Studies in first world countries involving mortality rather that life expectancy have also found a relationship with income level (Blakely et al. 2004, Kalwij et al. 2013, von Gaudecker & Scholz 2007).

Just 16% of the increase in life expectancy between 1930s and 1960s could be explained by rising income levels Preston (1975). Which seems to indicate that a coutries life expectancy is dependant on more than income levels.

Kalwij (2014)

Oeppen (2019) Very Good!!

Preston (1975) is a seminal work according to Oeppen (2019)

inequality Hu et al. (2015)

Chetty et al. (2016) in the US

income inequality does not affect health of a a country Jason Beckfield (2004)

Tarkiainen et al. (2012) (To be downloaded)

2.3.2 Education atainment

Kaplan et al. (2015) investigated the relationship between educational atainment and life expectancy in eight states in the United States. They found that even when controlling for variables like income, race, sex and common medical issues like cardiovascular disease, the relationship between educational antainment and life expectance remains statistically significant.

Luy et al. (2019) studied 3 developed nations, namely the United States, Italy and Denmark. They have also found a strong correlation between education levels and longevity.

But what is the nature of this correlation? According to Deary & Gottfredson (2004) Intelligence Quotient or IQ could explain the association. While Hayward et al. (2015) does not believe in a “causal relationship” but rather that it depends on factors like “time, place, and the social environment”.

In an attempt to find a causal relationship between education and life expectancy, van Kippersluis et al. (2011) investigated the result of the Netherlands increasing the mandatory number of years a child had to attend school to 7 years. It was 6 years previously. van Kippersluis et al. (2011) found a decrease in mortality of 3% for 81 year old males who had the additional year of schooling.

This relationship appears strongest in more developed countries where the life expectancy is already above 60 years (Bulled & Sosis 2010). In these countries, any educational investment leads to greater compensation for the learner than they would get in a less developed country Bulled & Sosis (2010), Handwerker (1986). In addition, Kabir (2008) also studied this relationship, among others, with regards to developing countries and did not find a correlation.

The question remains, which educational indicators should be used when investicating the relationship between education and life expectancy?

Various educational indicators have been used in the literature for comparing to life expectancy. One approach is to use the International Classification of Education (ISCED) system (UNESCO Institute for Statistics 2012). The ISCED 2011 standard consists of 9 levels ranging from ISCED level 0 (Early childhood education) to ISCED level 8 (Doctoral or equivalent level).

Luy et al. (2019) used the United Nations ISCED-97 (consisting of 7 levels) scale to break education attainment down into 3 levels namely Low (None to Lower Secondary), Medium (Upper secondary) and High (Tertiary education). In van Kippersluis et al. (2011) the Dutch SOI system (Standaard Onderwijs Indeling). Which according to van Kippersluis et al. (2011) is similar to the ISCED system. While in Deboosere et al. (2009) educational attainment was broken into 5 levels also ranging from no education to Tertiarty education.

Kaplan et al. (2015) broke educational attainment into 4 levels ranging from less than high school to college graduate.

In the study Bulled & Sosis (2010), the relationship between educational invesments and fertility against life expectancy, over 193 countries, was investigated. They used adult literacy and the enrolment ratios for primary,secondary and tertiary schooling.

For more see Montez & Friedman (2015) Much information!!!!

helping individuals to mobilise health resources Elo & Preston (1996) from Deboosere et al. (2009)

Study in Belgium Deboosere et al. (2009)

Inverse relationship Hoque et al. (2019)

netherlands van Kippersluis et al. (2011)

van Baal et al. (2016)

2.3.3 Per capita spending on health

Healthcare spending and life expectancy in the Unites States, between 1960 and 2000, was compared in Cutler et al. (2006). They found that the increased spending on health per capita, controlling for inflation, is positively correlated to US life expectancy for the time period in question.

Most Eastern European countries, who have joined the European Union, have seen an increase in healthcare spending. This has generally been accompanied by an increase in life expectancy (Jakovljevic et al. 2016). This has to be seen in the light of the so called "Russian Mortality Crisis" where former Soviet Union countries faced a sudden drop in life expectancy after the fall of the Berlin wall (Brainerd & Cutler 2005). Jakovljevic et al. (2016) found that the best metric to use when comparing health spending of coutries to be their total per capita health spending in US dollars.

The same relationship was found in Canada. When spending on healthcare is decreased, life expectancy follows (Crémieux et al. 1999).

It is well known that life expectancy in Sub-Saharan Africa is low. Here spending on health care can also be correlated to increases in life expectancy. Even though poor governance can undo some of the effects of increased spending (Makuta & O'Hare 2015).

A countries per capita healthcare is not neccesarily in proportion to its per capita income. In 2005 the United States spent 50% more on healthcare per capita than its income per capita would suggest (Anderson & Frogner 2008).

Shaw et al. (2005) showed that pharmaceutical expenditures shows a positive correlation with life expectancy in OECD countries.

medical spending Cutler et al. (2006)

?Grossman? 2017 determinants of health: an economic perspective ???? 1972 The Demand for Health: A Theoretical and Empirical Investigation,

Grossman (2000)

2.3.4 Unemployment

According to Bonamore et al. (2015), the literature has 2 main views on the relationship between unemployment and life expectancy. The first view states that during an economic downturn, people suffer from more stress and depression. This leads to more unhealthy lifestyle choices like smoking and alcohol. Which in turn lower life expectancy. Bonamore et al. (2015) cites the works of Lundin et al. (2014), Montgomery et al. (2013), Garcy & Vågerö (2012), Browning & Heinesen (2012), Dávalos et al. (2012), Backhans & Hemmingsson (2011), Deb et al. (2011) and Strully (2009), who take this view. The second view focusses on times when there is economic growth i.e. less unemployment. This period of economic growth also cal lead to stress eg. burning out and having less time for activities that benifit ones health. Like going to the gym. This view is held by Tapia Granados & Ionides (2011), Ruhm (2005), Tapia Granados (2005), Neumayer (2004) and Ruhm (2000). Then there are also studies express the view that no connection can be established (Bonamore et al. 2015). The view of Bonamore et al. (2015) is that the relationship is non-linear.

unemployment Bonamore et al. (2015) Roelfs et al. (2011) Roelfs et al. (2015)

2.4 Cross country studies

Bulled & Sosis (2010) Pearsons r and multivariate regression.Bulled & Sosis (2010) aims to show correlation not prediction.

Shaw et al. (2005) assumes a linear model and uses regression.

Kabir (2008) investigated how well the life expectancy of 93 developing countries were predicted by indicators like income,education and fertility (among others). It classied a countries life expectancy into 3 categories. Then used a probit model where the input variables have a linear relationship. Multiple Ordinary Least Squares Regression was then applied to study inicators' influences.

The study Hu et al. (2015), also used a linear regression model of GDP per capita, Gini indeces, ect with respect to life expectancy. The intention of the study was to link income inequality to mortality rates and life expectancy.

2.5 Multi variate studies

3 Methodology/Procedure

This study will attempt to create a model that can predict life expectancy for a country based on various socio-economic conditions in the country. Then evaluate how well various machine learning techniques model the relationships.

The philosophical standpoint of this study is Positivism. By using the scientific method, this study will comprise of an experiment to inductively determine whether machine learning techniques can provide more accurate life expectancy models than those created using regression. This cross-sectional study will use life expectancy indicators shown, from the literature, to have some correlation to life expectancy.

There are many studies that attempt to extrapolate future life expectancy for countries based on current data. This includes studies for high income countries (Kontis et al. 2017) and low income countries ????cite.

segment data into groups where each group has the same amount of data points???

Unlike Shaw et al. (2005), this study will not take into account the age distribution of each country.

As for HDI from Bulled & Sosis (2010) Adult literacy rate

primary secondary and tertiary enrolment ratios

GDP per Capita (Purchasing power parity)

"National datasets must be regarded with some level of caution as data gaps and issues of inconsistency and incoherence remain owing to differences in the effectiveness of infrastructure, political agendas, and additional factors, such as internal conflicts" Bulled & Sosis (2010)

The impact of finishing secondary school is different before vs after the second world war Deboosere et al. (2009)

3.1 Choice of dataset

3.1.1 Life Expectancy

For life expectancy data, this study will use the indicator "Life expectancy at birth, total(years) {SP.DYN.LE00.IN}" from the World Bank's Development Indicators Database (World Bank Group 2019*c*). This is a weighted average combining both male and female life expectancy and is calculated in a period life table (see Section 2.2). Only the data that is available for the last 30 years (1988-2017) will be used. This is in keeping with other studies where the amount of years that their studies look back on is limited to relatively recently (Luy et al. 2019, Hu et al. 2015, Tarkiainen et al. 2012, Kabir 2008, Shaw et al. 2005).

3.1.2 Income

This study will use GDP per capita as was used in Oeppen (2019), Shkolnikov et al. (2019), Mackenbach & Looman (2013) and De Vogli et al. (2005). The source will be GDP per capita, PPP (constant 2011 international $){NY.GDP.PCAP.PP.KD} from the World Bank (World Bank Group 2019*a*).

3.1.3 Educational Atainment

Just as in Bulled & Sosis (2010), this study will use adult literacy, and primary, secondary and tertiary enrolment ratios. For adult literacy indicator SE.ADT.LITR.ZS will be used from the world bank website (World Bank Group 2019*b*). The world bank credits UNESCO for the data. The indicator describes the percentage of adults, from the eage of 15, who can read and write to a certain level of proficiency.

3.2 Choise of algorithms

3.2.1 Algorithms Chosen

Regression Linear Regression is a popular technique, used to find relationships in data. As the name suggests Linear Regression assumes a linear relationship between the input variables and the result (Murphy 2012). This might not be the case for the target function. The target function could be any potential function. In the case of life expectancy modelling, we know that according to the Preston curve (Section 2.3.1) the relationship between income and life expectancy is not linear. Thus using Linear Regression should return a sub-optimal result. The same logic applies to Logistic Regression. It assumes a linear relationhip between inputs. The difference is that this linear sum is passed through the sigmoid function (Murphy 2012). This alsom makes it inappropriate for non-linear target functions. In this study Linear and Logistic Regression will be used as a baseline for comparison on the dataset.

Ridge Regression

k-Nearest Neighbour The k-Nearest Neighbour algorithm (kNN) is an instance based form of machine learning. It uses the classification of those datapoints closest to the datapoint to be classied to determine its classification. The kN-algorithm allows for non-linear problems spaces to be classified, because it does not make an assumption on the nature of the problem space. Additionally, how the algorithm determined its output value is transparent and can be used to study how various components affects the end result. In this study the standard kNN-algoritm will be altered to accomodate a real valued output and not just a class classification. This will be accomplised by taking the mean life expectancy for all the datapoints determined to be closest to the target point. Care will have to be taken to reduce the number of features the data, because this algorithm is sensitive to the so-called "curse of dimentionality" (Mitchell 1997).

using PCA to reduce dimentions
compare different distance metrics
can be used to determine missing data

Support Vector Machines The classic Support Vector Machine (SVM) is used in classification tasks. It involves determining the decision surface with regards to the datapoints closest to the surface. This study will use a modified SVM algorithm that makes it usefull for regression tasks. The SVN regression is described in Smola & Schölkopf (2004). The final model can be descriped as:

$$f(x) = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)k(x_i, x) + b \tag{1}$$

where α_i and α_i^* are Lagrange multipliers, b is the bias and the term $k(x_i, x)$ is the kernel. The kernel is a function to determine the similarity between 2 input vectors. Various kernels will be investigated, including the Radial-basis function

$$k(x_i, x) = exp(-\frac{|| x - x_i ||^2}{2\sigma^2}) \tag{2}$$

and the polynomial kernel

$$k(x_i, x) = (x_i.x + 1)^p \tag{3}$$

3.2.2 Ignored Algorithms

Neural Networks Even though Neural networks are capable of representing non-linear hypothesis spaces (Mitchell 1997), theyare not appropriate for this study for a couple of reasons. Firsly, the datasets that are available are not large enough. Neural networks typically require thousands if not tens of thousands of datapoints. Secondly the amount of processing power and processing time required, will not be available to this study. Thirdly, the results of nural networks are hard to interperet. How the Neural Network came to its conclusion is not clear to the researcher. Which makes it unsuitable as a tool to study the relationship between life expectancy and its various indicators.

Decision Trees Traceability and understandability are some of the hallmarks of Decision Trees. These algorithms are well suited problem spaces where the target function and the input attributes are discrete values. It is possible to approximate continuous input attributes by making a branch in the tree when a value is smaller or greter than some value or is between some value. For functions where input attributes span over large ranges, this leads to very large and sub-optimum trees (Mitchell 1997). The problem of determining life expectancy from socio-economic indices has a continuous target function output and continuous input attributes. Therefor, Decision Trees will be excluded from this study.

Clustering Clustering techniques are a form of unsupervised learning where the algorithm tries to group the data in a number of groups. Inessence, classifying the data into categories. This study will not attempt to classify the data, but rather find appropirate machine learning algorithms that could replce linear and logistic regression. While at the same time being able to predict the output of an input vector without needing to classify it first.

Bayes

3.3 Cross-validation

By using stratified k-fold cross validation, this study will aim to reduce the impact of the relative small dataset that will be analysed. This form of cross validation will ensure that when the validation set is chosen, no important datapoints are ignored for training. The data will be broken down randomly into k subsets of equal size. Each data subset will also contain equal amounts of datapoints with low and high life expectancies, so that no dataset is completely towards one end of the data range. One data subset is chosen to be the validation subset and the remaining $k-1$ subsets are combined into the training set. The model is then trained on the training dataset andt its performance is measured against the validation subset. This is done k times in order for each subset to be the validation subset. For each of the training runs the mean of the error will be calculated (Mitchell 1997, Murphy 2012). The value of k will be dependant on the final dataset.

Appendices

Table 1: Life table for the total population: United States, 2003 (Arias 2007)

Age	Probability of dying between ages x to x+1 q(x)	Number surviving to age x l(x)	Number dying between ages x to x+1 d(x)	Person-years lived between ages x to x+1 L(x)	Total number of person-years lived above age x T(x)	Expectation of life at age x e(x)
0-1	0.006865	100,000	687	99,394	7,743,016	77.4
1-2	0.000469	99,313	47	99,290	7,643,622	77.0
2-3	0.000337	99,267	33	99,250	7,544,332	76.0
3-4	0.000254	99,233	25	99,221	7,445,082	75.0
4-5	0.000194	99,208	19	99,199	7,345,861	74.0
5-6	0.000177	99,189	18	99,180	7,246,663	73.1
6-7	0.000160	99,171	16	99,163	7,147,482	72.1
7-8	0.000147	99,156	15	99,148	7,048,319	71.1
8-9	0.000132	99,141	13	99,134	6,949,171	70.1
9-10	0.000117	99,128	12	99,122	6,850,036	69.1
10-11	0.000109	99,116	11	99,111	6,750,914	68.1
11-12	0.000118	99,105	12	99,100	6,651,803	67.1
12-13	0.000157	99,094	16	99,086	6,552,704	66.1
13-14	0.000233	99,078	23	99,067	6,453,618	65.1
14-15	0.000339	99,055	34	99,038	6,354,551	64.2
15-16	0.000460	99,022	46	98,999	6,255,513	63.2
16-17	0.000577	98,976	57	98,947	6,156,514	62.2
17-18	0.000684	98,919	68	98,885	6,057,566	61.2
18-19	0.000769	98,851	76	98,813	5,958,681	60.3
19-20	0.000832	98,775	82	98,734	5,859,868	59.3
20-21	0.000894	98,693	88	98,649	5,761,134	58.4
21-22	0.000954	98,605	94	98,558	5,662,485	57.4
22-23	0.000990	98,511	98	98,462	5,563,928	56.5
23-24	0.000997	98,413	98	98,364	5,465,466	55.5
24-25	0.000982	98,315	97	98,267	5,367,101	54.6
25-26	0.000960	98,219	94	98,171	5,268,835	53.6
26-27	0.000942	98,124	92	98,078	5,170,663	52.7
27-28	0.000936	98,032	92	97,986	5,072,585	51.7
28-29	0.000947	97,940	93	97,894	4,974,599	50.8

29-30	0.000974	97,847	95	97,800	4,876,705	49.8
30-31	0.001008	97,752	98	97,703	4,778,906	48.9
31-32	0.001046	97,654	102	97,603	4,681,203	47.9
32-33	0.001097	97,551	107	97,498	4,583,600	47.0
33-34	0.001162	97,444	113	97,388	4,486,102	46.0
34-35	0.001244	97,331	121	97,271	4,388,715	45.1
35-36	0.001336	97,210	130	97,145	4,291,444	44.1
36-37	0.001441	97,080	140	97,010	4,194,299	43.2
37-38	0.001567	96,940	152	96,864	4,097,289	42.3
38-39	0.001714	96,788	166	96,705	4,000,424	41.3
39-40	0.001874	96,623	181	96,532	3,903,719	40.4
40-41	0.002038	96,442	197	96,343	3,807,187	39.5
41-42	0.002207	96,245	212	96,139	3,710,844	38.6
42-43	0.002389	96,033	229	95,918	3,614,705	37.6
43-44	0.002593	95,803	248	95,679	3,518,787	36.7
44-45	0.002819	95,555	269	95,420	3,423,108	35.8
45-46	0.003064	95,285	292	95,139	3,327,688	34.9
46-47	0.003322	94,993	316	94,836	3,232,548	34.0
47-48	0.003589	94,678	340	94,508	3,137,713	33.1
48-49	0.003863	94,338	364	94,156	3,043,205	32.3
49-50	0.004148	93,974	390	93,779	2,949,049	31.4
50-51	0.004458	93,584	417	93,375	2,855,270	30.5
51-52	0.004800	93,167	447	92,943	2,761,895	29.6
52-53	0.005165	92,719	479	92,480	2,668,952	28.8
53-54	0.005554	92,241	512	91,984	2,576,472	27.9
54-55	0.005971	91,728	548	91,454	2,484,487	27.1
55-56	0.006423	91,181	586	90,888	2,393,033	26.2
56-57	0.006925	90,595	627	90,281	2,302,145	25.4
57-58	0.007496	89,968	674	89,630	2,211,864	24.6
58-59	0.008160	89,293	729	88,929	2,122,234	23.8
59-60	0.008927	88,565	791	88,169	2,033,305	23.0
60-61	0.009827	87,774	863	87,343	1,945,136	22.2
61-62	0.010831	86,911	941	86,441	1,857,793	21.4
62-63	0.011872	85,970	1021	85,460	1,771,352	20.6
63-64	0.012891	84,949	1095	84,402	1,685,892	19.8
64-65	0.013908	83,854	1166	83,271	1,601,490	19.1
65-66	0.015003	82,688	1241	82,068	1,518,219	18.4
66-67	0.016267	81,448	1325	80,785	1,436,151	17.6
67-68	0.017699	80,123	1418	79,414	1,355,366	16.9
68-69	0.019320	78,705	1521	77,944	1,275,953	16.2
69-70	0.021108	77,184	1629	76,369	1,198,008	15.5
70-71	0.022950	75,555	1734	74,688	1,121,639	14.8
71-72	0.024904	73,821	1838	72,902	1,046,951	14.2
72-73	0.027151	71,982	1954	71,005	974,050	13.5
73-74	0.029784	70,028	2086	68,985	903,044	12.9
74-75	0.032753	67,942	2225	66,830	834,059	12.3
75-76	0.035831	65,717	2355	64,540	767,230	11.7
76-77	0.038987	63,362	2470	62,127	702,690	11.1
77-78	0.042503	60,892	2588	59,598	640,563	10.5
78-79	0.046557	58,304	2714	56,947	580,965	10.0

79-80	0.051200	55,589	2846	54,166	524,019	9.4
80-81	0.056335	52,743	2971	51,258	469,853	8.9
81-82	0.061837	49,772	3078	48,233	418,595	8.4
82-83	0.067856	46,694	3168	45,110	370,362	7.9
83-84	0.074504	43,526	3243	41,904	325,252	7.5
84-85	0.081975	40,283	3302	38,632	283,348	7.0
85-86	0.089682	36,981	3317	35,322	244,716	6.6
86-87	0.098031	33,664	3300	32,014	209,394	6.2
87-88	0.107059	30,364	3251	28,739	177,380	5.8
88-89	0.116804	27,113	3167	25,530	148,641	5.5
89-90	0.127300	23,946	3048	22,422	123,111	5.1
90-91	0.138581	20,898	2896	19,450	100,689	4.8
91-92	0.150676	18,002	2712	16,646	81,239	4.5
92-93	0.163611	15,289	2502	14,039	64,594	4.2
93-94	0.177408	12,788	2269	11,654	50,555	4.0
94-95	0.192080	10,519	2021	9,509	38,901	3.7
95-96	0.207636	8,499	1765	7,616	29,392	3.5
96-97	0.224075	6,734	1509	5,980	21,776	3.2
97-98	0.241387	5,225	1261	4,594	15,796	3.0
98-99	0.259552	3,964	1029	3,449	11,202	2.8
99-100	0.278539	2,935	818	2,526	7,752	2.6
100+	1.00000	2,118	2118	5,226	5,226	2.5

References

Anderson, G. F. & Frogner, B. K. (2008), 'Health spending in OECD countries: Obtaining value per dollar', *Health Affairs* **27**(6), 1718–1727.

Arias, E. (2007), National Vital Statistics Reports, Technical Report 6.

Arias, E., Heron, M. & Xu, J. (2017), 'United States Life Tables, 2013.', *National vital statistics reports : from the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System* **66**(3), 1–64.

Backhans, M. C. & Hemmingsson, T. (2011), 'Unemployment and mental healthwho is (not) affected?', *The European Journal of Public Health* **22**(3), 429–433.

Blakely, T., Kawachi, I., Atkinson, J. & Fawcett, J. (2004), 'Income and mortality: The shape of the association and confounding New Zealand Census - Mortality study, 1981-1999', *International Journal of Epidemiology* **33**(4), 874–883.

Bonamore, G., Carmignani, F. & Colombo, E. (2015), 'Addressing the unemployment-mortality conundrum: Non-linearity is the answer', *Social Science and Medicine* **126**, 67–72.

Brainerd, E. & Cutler, D. M. (2005), Autopsy on an Empire: Understanding Mortality in Russia and the Former Soviet Union, Technical Report 1.

Browning, M. & Heinesen, E. (2012), 'Effect of job loss due to plant closure on mortality and hospitalization', *Journal of Health Economics* **31**(4), 599–616.

Bulled, N. L. & Sosis, R. (2010), 'Examining the Relationship between Life Expectancy, Reproduction, and Educational Attainment', *Human Nature* **21**(3), 269–289.

Chen, J. H. & Asch, S. M. (2017), 'Machine Learning and Prediction in Medicine Beyond the Peak of Inflated Expectations', *New England Journal of Medicine* **376**(26), 2507–2509.

Chetty, R., Stepner, M., Abraham, S., Lin, S., Scuderi, B., Turner, N., Bergeron, A. & Cutler, D. (2016), 'The association between income and life expectancy in the United States, 2001-2014', *JAMA - Journal of the American Medical Association* **315**(16), 1750–1766.

Crémieux, P.-Y., Ouellette, P. & Pilon, C. (1999), 'Health care spending as determinants of health outcomes', *Health Economics* **8**(7), 627–639.

Cutler, D. M., Rosen, A. B. & Vijan, S. (2006), 'The Value of Medical Spending in the United States, 19602000'.

Dávalos, M. E., Fang, H. & French, M. T. (2012), 'EASING THE PAIN OF AN ECONOMIC DOWNTURN: MACROECONOMIC CONDITIONS AND EXCESSIVE ALCOHOL CONSUMPTION', *Health Economics* **21**(11), 1318–1335.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/hec.1788*

De Vogli, R., Mistry, R., Gnesolto, R. & Cornia, G. A. (2005), 'Has the relation between income inequality and life expectancy disappeared? Evidence from Italy and top industrialised countries', *Journal of Epidemiology and Community Health* **59**(2), 158–162.
URL: *http://jech.bmj.com/*

Deary, I. J. & Gottfredson, L. S. (2004), 'Intelligence Predicts Health and Longevity, but Why?', *Current Directions in Psychological Science* **13**(1), 1–4.

Deb, P., Gallo, W. T., Ayyagari, P., Fletcher, J. M. & Sindelar, J. L. (2011), 'The effect of job loss on overweight and drinking', *Journal of Health Economics* **30**(2), 317–327.

Deboosere, P., Gadeyne, S. & Van Oyen, H. (2009), 'The 1991-2004 evolution in life expectancy by educational level in Belgium based on linked census and Population register data', *European Journal of Population* **25**(2), 175–196.

Elo, I. T. & Preston, S. H. (1996), 'Educational differentials in mortality: United States, 1979-85', *Social Science and Medicine* **42**(1), 47–57.

Garcy, A. M. & Vågerö, D. (2012), 'The length of unemployment predicts mortality, differently in men and women, and by cause of death: A six year mortality follow-up of the Swedish 19921996 recession', *Social Science & Medicine* **74**(12), 1911–1920.

Gompertz, B. (1825), 'XXIV. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. In a letter to Francis Baily, Esq. F. R. S. &c', *Philosophical Transactions of the Royal Society of London* **115**, 513–583.

Grossman, M. (2000), THE HUMAN CAPITAL MODEL, *in* 'Handbook of Health Economics, Volume 1', Vol. 1, pp. 348–407.

Handwerker, W. P. (1986), 'The Modern Demographic Transition: An Analysis of Subsistence Choices and Reproductive Consequences', *American Anthropologist* **88**(2), 400–417.

Hayward, M. D., Hummer, R. A. & Sasson, I. (2015), 'Trends and Group Differences in the Association between Educational Attainment and U.S. Adult Mortality: Implications for Understanding Education's Causal Influence *', *Soc Sci Med* **127**, 8–18.

Hoque, M. M., King, E. M., Montenegro, C. E. & Orazem, P. F. (2019), 'Revisiting the relationship between longevity and lifetime education: global evidence from 919 surveys', *Journal of Population Economics* **32**(2), 551–589.

Hu, Y., van Lenthe, F. J. & Mackenbach, J. P. (2015), 'Income inequality, life expectancy and cause-specific mortality in 43 European countries, 19872008: a fixed effects study', *European Journal of Epidemiology* **30**(8), 615–625.

Jakovljevic, M. B., Vukovic, M. & Fontanesi, J. (2016), 'Life expectancy and health expenditure evolution in Eastern EuropeDiD and DEA analysis', *Expert Review of Pharmacoeconomics and Outcomes Research* **16**(4), 537–546.

Jason Beckfield (2004), 'Does Income Inequality Harm Health? New Cross-National Evidence', *Journal of Health and Social Behavior* **45**(3), 231–248.

Kabir, M. (2008), 'Determinants of Life Expectancy in Developing Countries', *The Journal of Developing Areas* **41**(2), 185–204.

Kalwij, A. S. (2014), 'An empirical analysis of the importance of controlling for unobserved heterogeneity when estimating the income-mortality gradient', *Demographic Research* **31**(1), 913–940.

Kalwij, A. S., Alessie, R. J. M., Knoef, M. G., Kalwij, A. S., Alessie, R. J. M. & Knoef, M. G. (2013), 'The Association Between Individual Income and Remaining Life Expectancy at the Age of 65 in the Netherlands', *Demography* **50**(1), 181–206.

Kaplan, R. M., Howard, V. J., Safford, M. M. & Howard, G. (2015), 'Educational attainment and longevity: Results from the REGARDS U.S. national cohort study of blacks and whites', *Annals of Epidemiology* **25**(5), 323–328.

Kontis, V., Bennett, J. E., Mathers, C. D., Li, G., Foreman, K. & Ezzati, M. (2017), 'Future life expectancy in 35 industrialised countries: projections with a Bayesian model ensemble', *The Lancet* **389**(10076), 1323–1335.

Lundin, A., Falkstedt, D., Lundberg, I. & Hemmingsson, T. (2014), 'Unemployment and coronary heart disease among middle-aged men in Sweden: 39 243 men followed for 8years', *Occupational and Environmental Medicine* **71**(3), 183 LP – 188.

Luy, M., Zannella, M., Wegner-Siegmundt, C., Minagawa, Y., Lutz, W. & Caselli, G. (2019), 'The impact of increasing education levels on rising life expectancy: a decomposition analysis for Italy, Denmark, and the USA', *Genus* **75**(1), 11.

Mackenbach, J. P. & Looman, C. W. (2013), 'Life expectancy and national income in Europe, 1900-2008: An update of Preston's analysis', *International Journal of Epidemiology* **42**(4), 1100–1110.
URL: *http://www.ggdc.net/MADDISON/oriindex.*

Makeham, W. M. (1860), 'On the Law of Mortality and Construction of Annuity Tables', *The Assurance Magazine and Journal of the Institute of Actuaries* **8**(06), 301–310.

Makuta, I. & O'Hare, B. (2015), 'Quality of governance, public spending on health and health status in Sub Saharan Africa: a panel data regression analysis', *BMC Public Health* **15**(1), 932.

Mitchell, T. M. (1997), *Machine Learning*, McGraw-Hill international editions - computer science series, McGraw-Hill.

Montez, J. K. & Friedman, E. M. (2015), 'Educational attainment and adult health: Under what conditions is the association causal?', *Social Science and Medicine* **127**(1), 1–7.

Montgomery, S., Udumyan, R., Magnuson, A., Osika, W., Sundin, P.-O. & Blane, D. (2013), 'Mortality following unemployment during an economic downturn: Swedish register-based cohort study', *BMJ Open* **3**(7), e003031.

Murphy, K. P. (2012), *Machine Learning: A Probabilistic Perspective.*

Neumayer, E. (2004), 'Recessions lower (some) mortality rates:: evidence from Germany', *Social Science & Medicine* **58**(6), 1037–1047.

Oeppen, J. (2019), Life Expectancy Convergence Among Nations Since 1820: Separating the Effects of Technology and Income, *in* T. Bengtsson & N. Keilman, eds, 'Old and New Perspectives on Mortality Forecasting', Springer International Publishing, Cham, pp. 197–219.

Preston, S. H. (1975), 'The Changing Relation between Mortality and level of Economic Development', *Population Studies* **29**(2), 231–248.

Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Liu, P. J., Liu, X., Sun, M., Sundberg, P., Yee, H., Zhang, K., Duggan, G. E., Flores, G., Hardt, M., Irvine, J., Le, Q., Litsch, K., Marcus, J., Mossin, A., Tansuwan, J., Wang, D., Wexler, J., Wilson, J., Ludwig, D., Volchenboum, S. L., Chou, K., Pearson, M., Madabushi, S., Shah, N. H., Butte, A. J., Howell, M., Cui, C., Corrado, G. & Dean, J. (2018), 'Scalable and accurate deep learning for electronic health records', *Digital Medicine* **1**, 18.

Roelfs, D. J., Shor, E., Blank, A. & Schwartz, J. E. (2015), 'Misery loves company? A meta-regression examining aggregate unemployment rates and the unemployment-mortality association', *Annals of Epidemiology* **25**(5), 312–322.

Roelfs, D. J., Shor, E., Davidson, K. W. & Schwartz, J. E. (2011), 'Losing life and livelihood: A systematic review and meta-analysis of unemployment and all-cause mortality', *Social Science and Medicine* **72**(6), 840–854.

Ruhm, C. (2000), 'Are Recessions Good for Your Health?', *The Quarterly Journal of Economics* **115**(2), 617–650.

Ruhm, C. J. (2005), 'Healthy living in hard times', *Journal of Health Economics* **24**(2), 341–363.

Shaw, J. W., Horrace, W. C. & Vogel, R. J. (2005), The Determinants of Life Expectancy: An Analysis of the OECD, Technical Report 4.

Shkolnikov, V. M., Andreev, E. M., Tursun-zade, R. & Leon, D. A. (2019), 'Patterns in the relationship between life expectancy and gross domestic product in Russia in 200515: a cross-sectional analysis', *The Lancet Public Health* **4**(4), e181–e188.

Smola, A. J. & Schölkopf, B. (2004), 'A tutorial on support vector regression', *Statistics and Computing* **14**(3), 199–222.

Strully, K. W. (2009), 'Job loss and health in the U.S. labor market.', *Demography* **46**(2), 221–246.

Tapia Granados, J. A. (2005), 'Increasing mortality during the expansions of the US economy, 1900-1996.', *International journal of epidemiology* **34**(6), 1194–1202.

Tapia Granados, J. A. & Ionides, E. L. (2011), 'Mortality and Macroeconomic Fluctuations in Contemporary Sweden', *European Journal of Population / Revue européenne de Démographie* **27**(2), 157–184.

Tarkiainen, L., Martikainen, P., Laaksonen, M. & Valkonen, T. (2012), 'Trends in life expectancy by income from 1988 to 2007: decomposition by age and cause of death', *Journal of Epidemiology and Community Health* **66**(7), 573 LP – 578.

UNESCO Institute for Statistics (2012), *International Standard Classification of Education ISCED 2011*.

van Baal, P., Peters, F., Mackenbach, J. & Nusselder, W. (2016), 'Forecasting differences in life expectancy by education', *Population Studies* **70**(2), 201–216.

van Kippersluis, H., O'Donnell, O. & van Doorslaer, E. (2011), 'Long Run Returns to Education: Does Schooling Lead to an Extended Old Age?', *Journal of Human Resources* **46**(4), 695–721.

von Gaudecker, H. M. & Scholz, R. D. (2007), 'Differential mortality by lifetime earnings in Germany', *Demographic Research* **17**, 83–108.

World Bank Group (2019*a*), 'GDP per capita, PPP (constant 2011 international $)'.
URL: *https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.KD*

World Bank Group (2019*b*), 'Literacy rate, adult total (% of people ages 15 and above)'.
URL: *https://data.worldbank.org/indicator/SE.ADT.LITR.ZS*

World Bank Group (2019*c*), 'World Development Indicators'.
URL: *http://datatopics.worldbank.org/world-development-indicators/*