How to live longer ?

Ashraf Fawaz, Pierre-Antoine Chabbert, Stéphane Salim

ENAC

April 25, 2022

Who wants to live forever ?

Figure: Queen's song *who wants to live forever*

Data and Hypothesises

- Data for the 50 US States, in years 2019, 2010 and 2000
- Harvested on America's Health Rankings and CDC websites

Hypothesis 1

The richest states have a higher life expectancy

Hypothesis 2

States with the healthiest lifestyle have a higher life expectancy

Hypothesis 3

Sunniest states have a higher life expectancy

Life expectancy by state

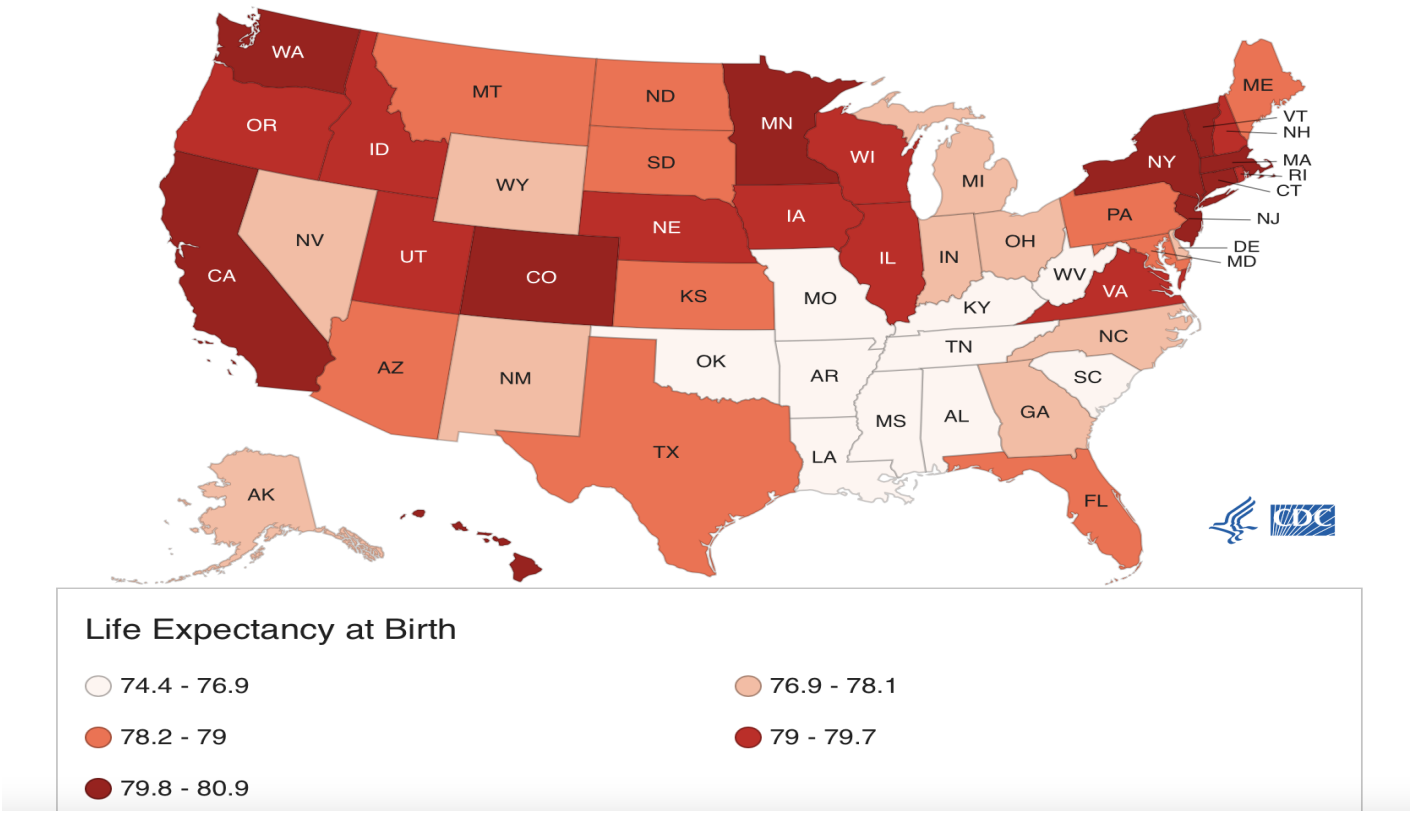


Figure: life expectancy by state in 2019

Descriptive statistics

Data	**Highest value**	**Mean value**	**Lowest score**
GDP	California (3 052 645)	421 962,5	Vermont(34 127,5)
Smoking	West Virginia (25,2%)	16,65%	Utah (9%)
med	Rhode Island (274,9)	156,5	Idaho(96,6)
Drinking	North Dakota(22,2%)	16,7%	Utah(11,2%)
over weight	Mississippi(40,8%)	32%	Colorado(23,8%)
nb sunny day	Arizona(193)	103	VT and WA(58)
air pol($\mu g/m^3$)	California(12,6)	7,276	New Hampshire(4,1)
Inactivity	Indiana(30,9%)	27,5%	Arizona(24,1%)

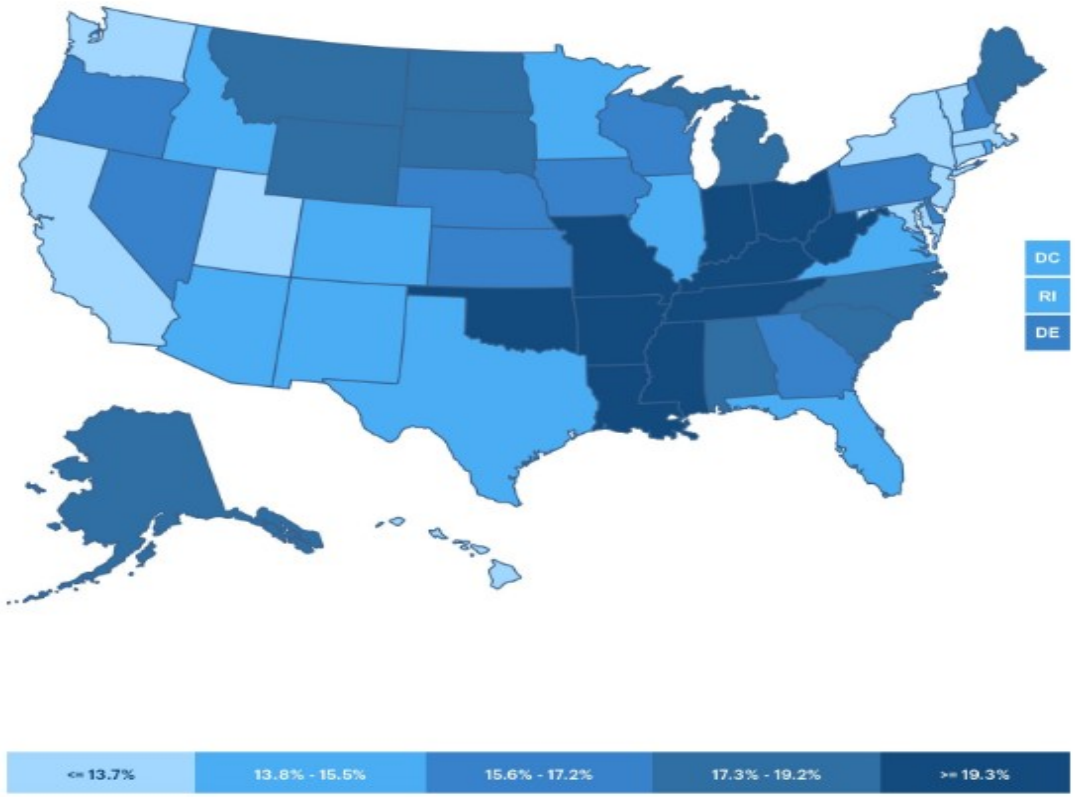
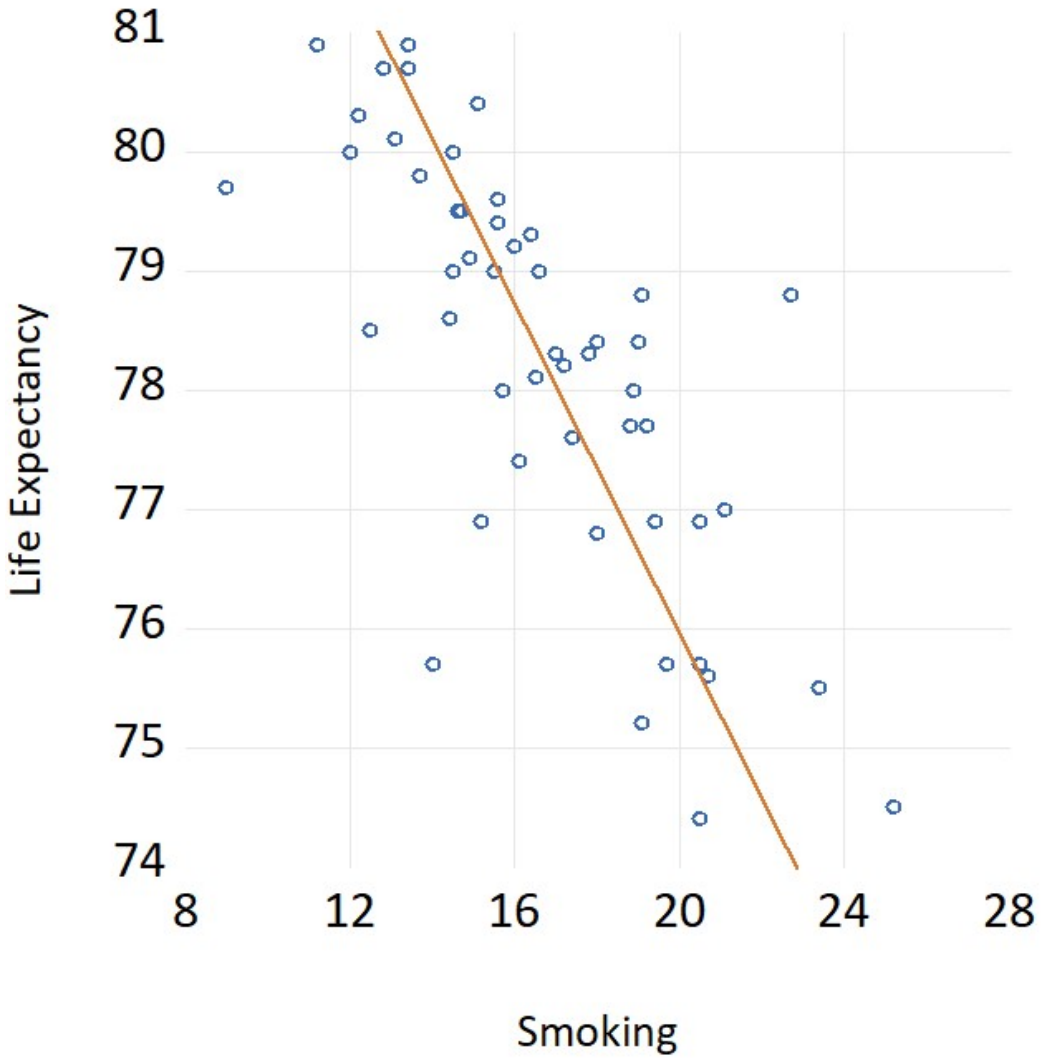


Figure: Percentage of adults who reported smoking at least 100 cigarettes in their lifetime and currently smoke daily or some days


81
80
79
78
77
76
75
74
Life Expectancy
8
12
16
20
24
28
Smoking

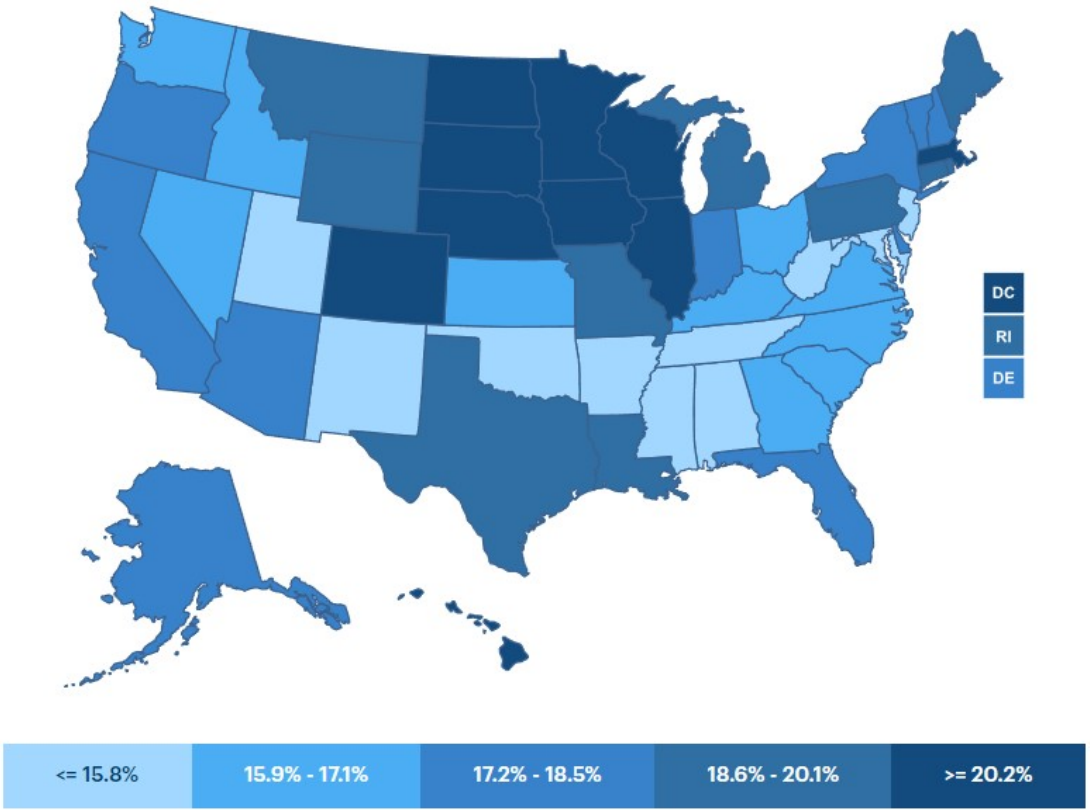


Figure: Percentage of adults who reported binge drinking (four or more [females] or five or more [males] drinks on one occasion in the past 30 days) or heavy drinking (eight or more [females] or 15 or more [males] drinks per week)

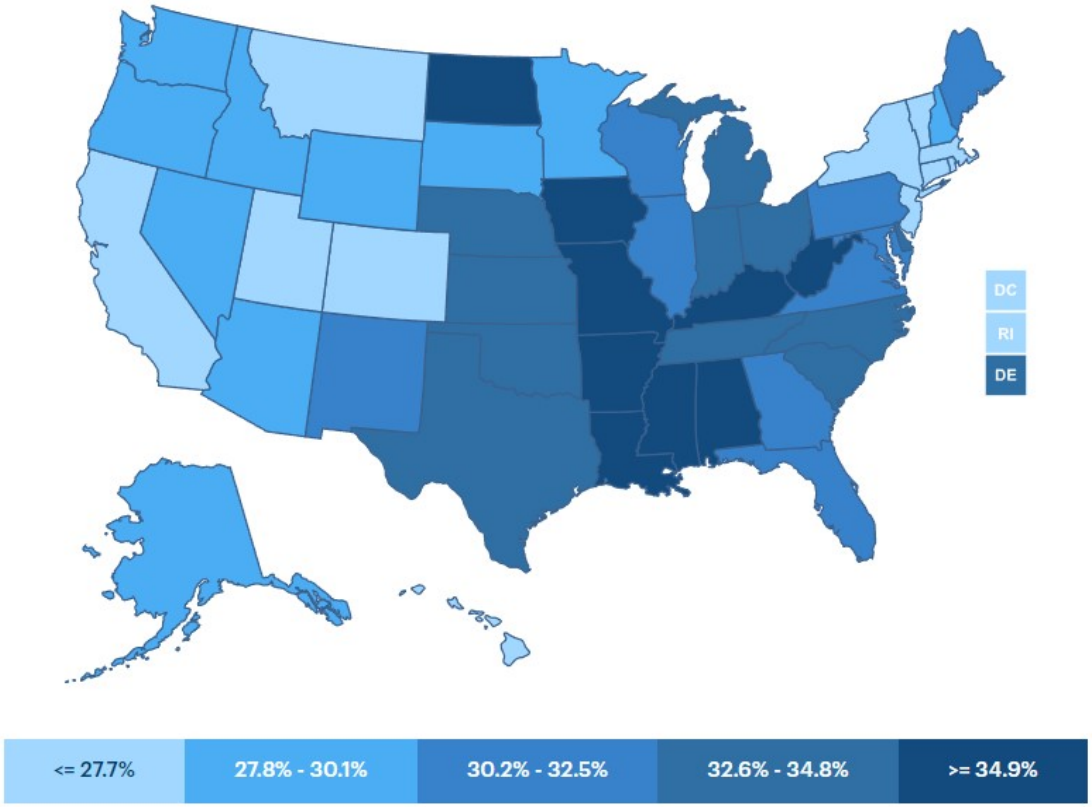


Figure: Percentage of adults with a body mass index of 30.0 or higher based on reported height and weight

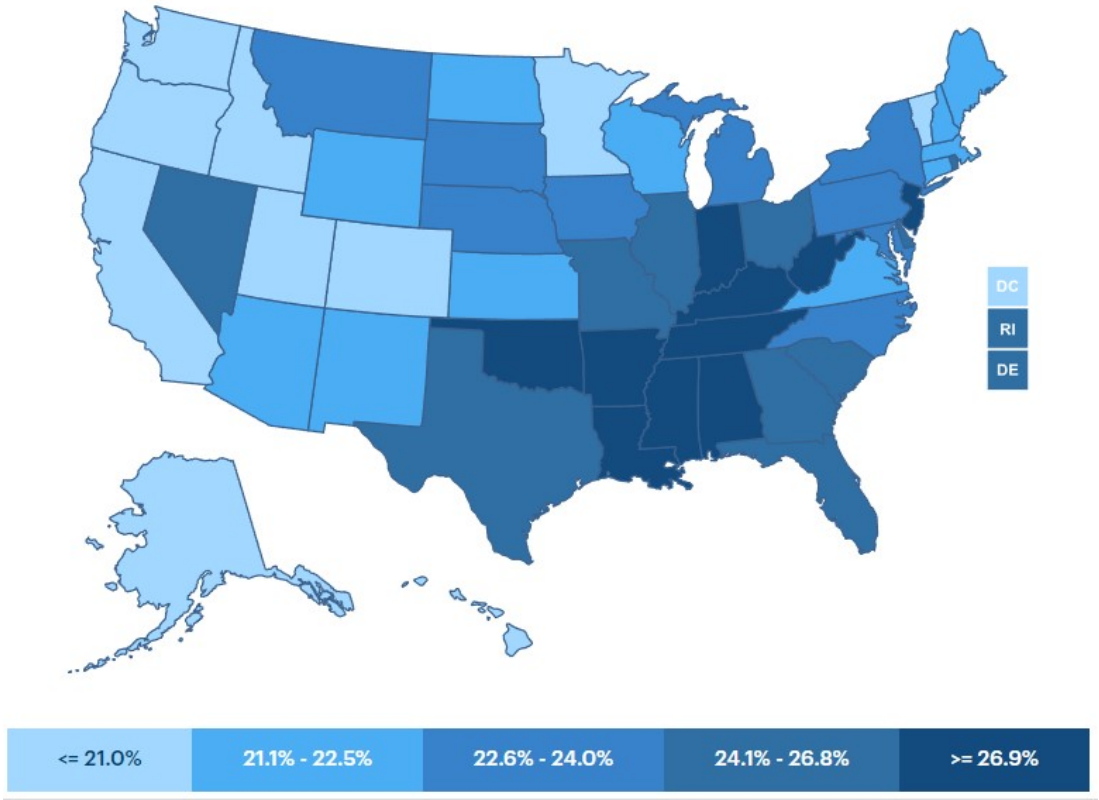
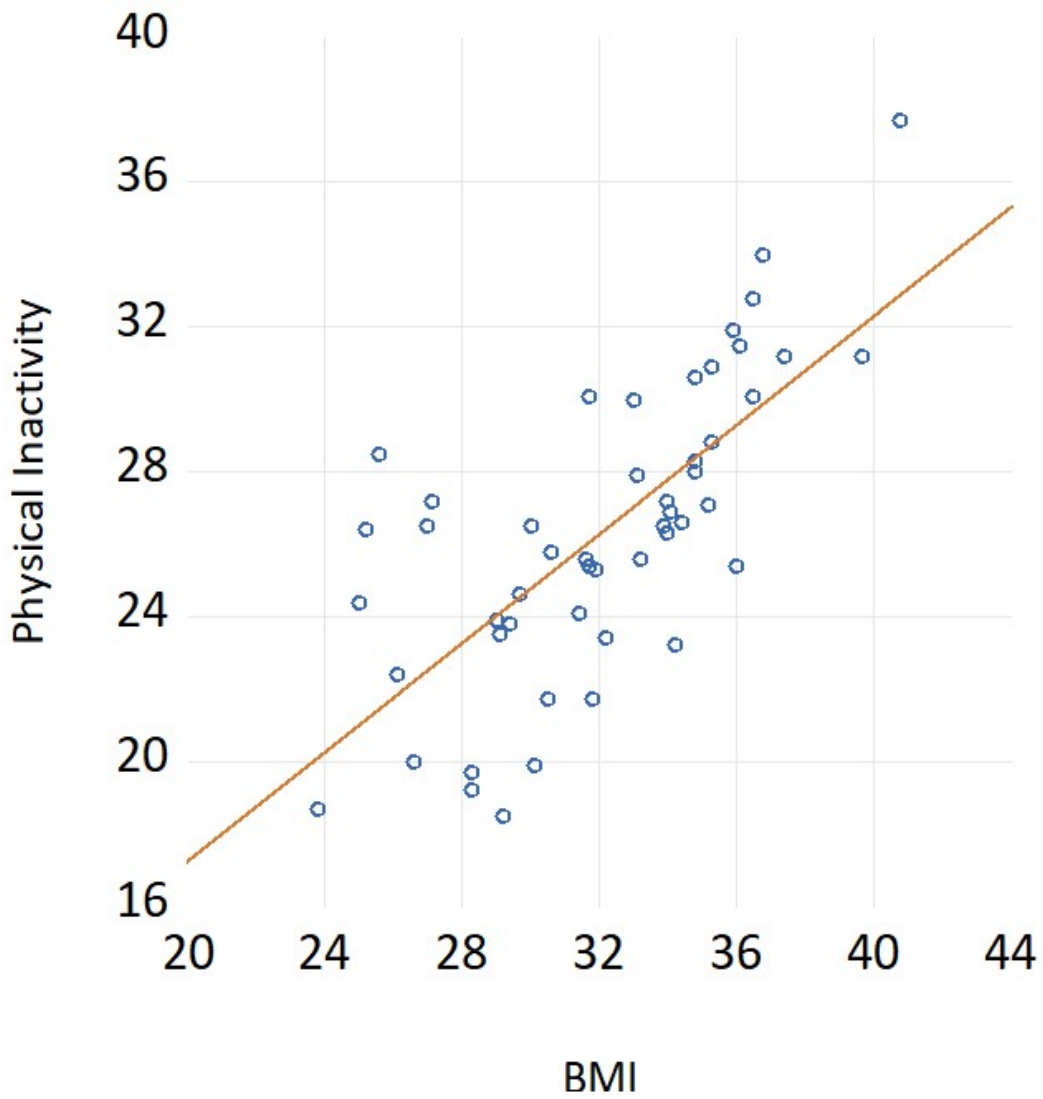
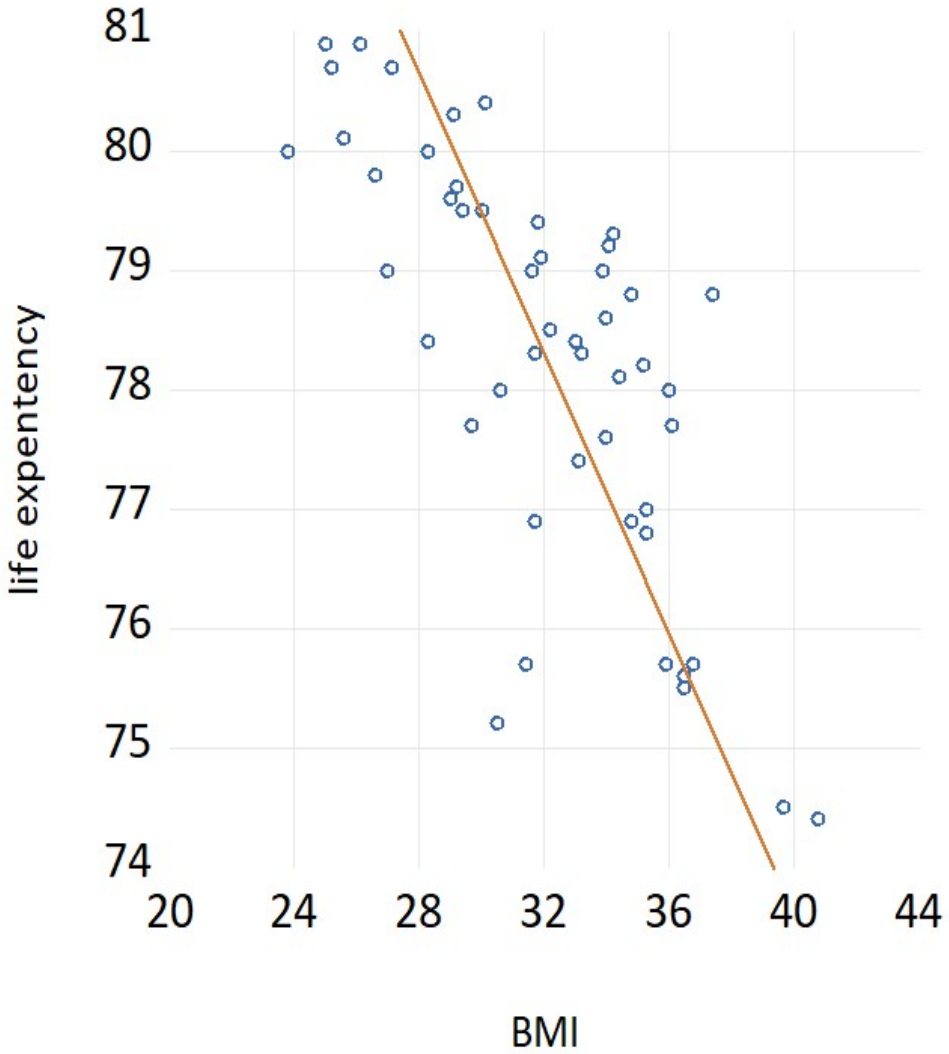
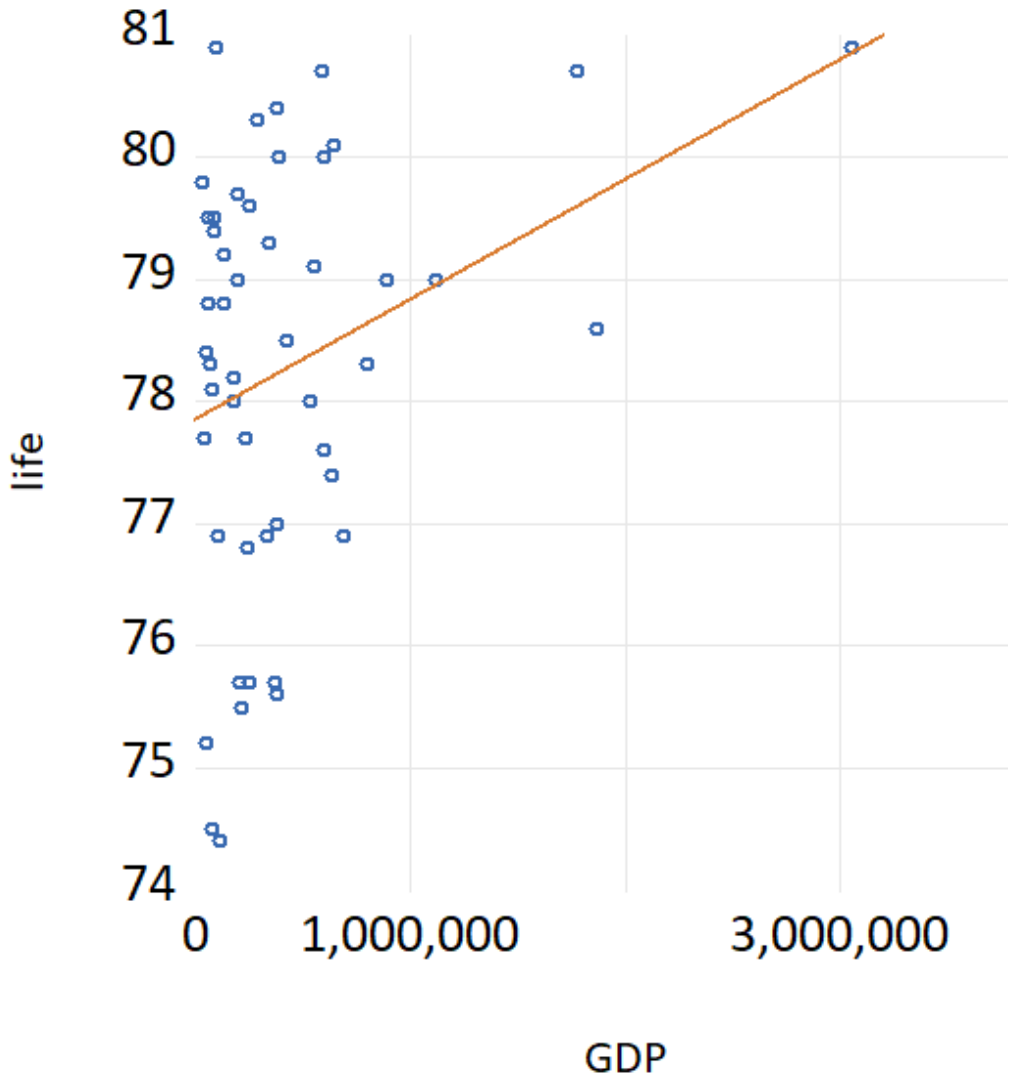


Figure: Percentage of adults who reported doing no physical activity or exercise other than their regular job in the past 30 days


40
36
32
28
24
20
16
Physical Inactivity
20
24
28
32
36
40
44
BMI


81
80
79
78
77
76
75
74
life expentency
20
24
28
32
36
40
44
BMI


81
80
79
78
77
76
75
74
life
0
1,000,000
3,000,000
GDP

Equation 01

- We want to model life expectancy across the different states
- The equation of our model can be written as :

$$life = \beta_0 + \beta_1 gdp * 10^{-7} + \beta_2 cig + \beta_3 drink + \beta_5 obesity + \beta_6 med + \beta_7 sun + \beta_8 air + \beta_9 inac \quad \text{(E1)}$$

First regression : 2019

Dependent Variable: LIFE
Method: Least Squares
Date: 04/23/22 Time: 11:37
Sample: 1 50
Included observations: 50

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	84.77736	2.086976	40.62210	0.0000
GDP*10^(-7)	3.718542	3.654006	1.017662	0.3148
CIG	-0.298394	0.082457	-3.618781	0.0008
DRINK	0.156719	0.057350	2.732660	0.0092
OBESITY	-0.100781	0.063834	-1.578806	0.1221
MED	0.001556	0.004507	0.345190	0.7317
SUN	-0.010428	0.006235	-1.672371	0.1021
AIR	-0.117225	0.134418	-0.872097	0.3882
INAC	0.022916	0.056487	0.405686	0.6871

R-squared	0.740513	Mean dependent var	78.27400
Adjusted R-squared	0.689881	S.D. dependent var	1.712715
S.E. of regression	0.953782	Akaike info criterion	2.904786
Sum squared resid	37.29772	Schwarz criterion	3.248950
Log likelihood	-63.61966	Hannan-Quinn criter.	3.035846
F-statistic	14.62548	Durbin-Watson stat	2.138598
Prob(F-statistic)	0.000000		

Figure: Eq01 output for 2019

First regression : 2010

Dependent variable: LIFE
Method: Least Squares
Date: 04/23/22 Time: 11:40
Sample (adjusted): 1 50
Included observations: 50 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	84.76046	1.375318	61.62970	0.0000
GDP*10^(-7)	2.603469	2.497063	1.042612	0.3032
CIG	-0.226087	0.035945	-6.289886	0.0000
DRINK	0.127895	0.025944	4.929651	0.0000
OBESITY	-0.088982	0.043919	-2.026064	0.0493
MED	0.005141	0.003644	1.410616	0.1659
SUN	-0.003601	0.003192	-1.127951	0.2659
AIR	-0.006666	0.047430	-0.140548	0.8889
INAC	-0.077000	0.039928	-1.928476	0.0607

R-squared	0.927244	Mean dependent var	78.55400
Adjusted R-squared	0.913047	S.D. dependent var	1.631728
S.E. of regression	0.481159	Akaike info criterion	1.536313
Sum squared resid	9.492089	Schwarz criterion	1.880477
Log likelihood	-29.40782	Hannan-Quinn criter.	1.667372
F-statistic	65.31566	Durbin-Watson stat	1.994644
Prob(F-statistic)	0.000000		

Figure: Eq01 output for 2010

First regression : 2000

Dependent Variable: LIFE
Method: Least Squares
Date: 04/23/22 Time: 11:26
Sample (adjusted): 1 50
Included observations: 50 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	84.99268	1.509687	56.29819	0.0000
GDP*10^(-7)	0.259226	5.178626	0.050057	0.9603
CIG	-0.180654	0.053430	-3.381110	0.0015
DRINK	0.129312	0.040215	3.215513	0.0025
OBESITY	-0.239124	0.062137	-3.848310	0.0004
SUN	-0.004790	0.004492	-1.066516	0.2921
INAC	-0.021824	0.034460	-0.633310	0.5299

R-squared	0.711663	Mean dependent var	76.87200
Adjusted R-squared	0.671430	S.D. dependent var	1.452379
S.E. of regression	0.832519	Akaike info criterion	2.600455
Sum squared resid	29.80276	Schwarz criterion	2.868138
Log likelihood	-58.01138	Hannan-Quinn criter.	2.702391
F-statistic	17.68849	Durbin-Watson stat	2.572830
Prob(F-statistic)	0.000000		

Figure: Eq01 output for 2000

Robustness

Variables	**Units**	**2019**	**2010**	**2000**
GDP	Current million USD	+	+	+
CIG	Percentage smoking adults	-	-	-
DRINK	Percentage binge-drinking	+	+	+
OBESITY	Percentage $BMI > 30$	-	-	-
MED	Number per 100,000	+	+	NA
SUN	Number of clear days	-	+	-
AIR	Exposure in $\mu g/m^3$	-	+	NA
INAC	Percentage inac last 30 days	+	+	-

Table: Signs of the variables' coefficients

Robustness

- The coefficient of SUN has opposite signs from a year to another
- Let's add a new variable, SUN^2, and test whether or not this new model is better
- $life = \beta_0 + \beta_1 gdp * 10^{-7} + \beta_2 cig + \beta_3 drink + \beta_5 obesity + \beta_6 med + \beta_7 sun + \beta_8 sun^2 + \beta_9 air + \beta_{10} inac$

Second regression : 2019

Dependent Variable: LIFE
Method: Least Squares
Date: 04/23/22 Time: 13:19
Sample: 1 50
Included observations: 50

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	81.04376	2.186425	37.06680	0.0000
GDP*10^(-7)	3.774097	3.278331	1.151225	0.2565
CIG	-0.235444	0.076388	-3.082205	0.0037
DRINK	0.110179	0.053343	2.065467	0.0454
OBESITY	-0.111910	0.057369	-1.950692	0.0581
MED	0.004466	0.004138	1.079333	0.2869
SUN	0.076759	0.026951	2.848075	0.0069
SUN^2	-0.000370	0.000112	-3.307017	0.0020
AIR	-0.073286	0.121326	-0.604042	0.5492
INAC	-0.042906	0.054447	-0.788027	0.4353

R-squared	0.796226	Mean dependent var	78.27400
Adjusted R-squared	0.750377	S.D. dependent var	1.712715
S.E. of regression	0.855711	Akaike info criterion	2.703089
Sum squared resid	29.28966	Schwarz criterion	3.085493
Log likelihood	-57.57722	Hannan-Quinn criter.	2.848711
F-statistic	17.36624	Durbin-Watson stat	1.995720
Prob(F-statistic)	0.000000		

Figure: Eq02 output for 2019

F test

- We assume that all CAs hold
- We want to test if the general model (Eq02) can be nested within the restricted model (Eq01)
- $H_0 : R\beta = r, H_1 : /H_0$
  $F = \left(\frac{\widehat{u'_R}\widehat{u_R}}{\widehat{u'}\widehat{u}} - 1 \right) \left(\frac{n-k}{q} \right) \overset{H_0}{\sim} F(q, n-k)$
- if CA5 fails : $qF \overset{H_0}{\sim} \chi^2(q)$ *as* $n \to \infty$

F test

Heteroskedasticity Test: Breusch-Pagan-Godfrey
Null hypothesis: Homoskedasticity

F-statistic	1.005003	Prob. F(9,40)	0.4521
Obs*R-squared	9.221148	Prob. Chi-Square(9)	0.4171
Scaled explained SS	21.24619	Prob. Chi-Square(9)	0.0116

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 04/23/22 Time: 20:00
Sample: 1 50
Included observations: 50

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.961223	4.055199	0.237035	0.8138
GDP*10^(-7)	0.544378	6.080377	0.089530	0.9291
CIG	0.280574	0.141678	1.980358	0.0546
DRINK	-0.156442	0.098937	-1.581232	0.1217
OBESITY	-0.048239	0.106404	-0.453363	0.6527
MED	-0.000929	0.007674	-0.121076	0.9042
SUN	0.029503	0.049987	0.590205	0.5584
SUN^2	-0.000116	0.000207	-0.558840	0.5794
AIR	0.008710	0.225026	0.038707	0.9693
INAC	-0.096736	0.100984	-0.957937	0.3438

R-squared	0.184423	Mean dependent var	0.585793
Adjusted R-squared	0.000918	S.D. dependent var	1.587831
S.E. of regression	1.587102	Akaike info criterion	3.938553
Sum squared resid	100.7557	Schwarz criterion	4.320958
Log likelihood	-88.46383	Hannan-Quinn criter.	4.084175
F-statistic	1.005003	Durbin-Watson stat	2.108650
Prob(F-statistic)	0.452107		

Figure: Eq02 : BPG test for heteroscedasticity

F test

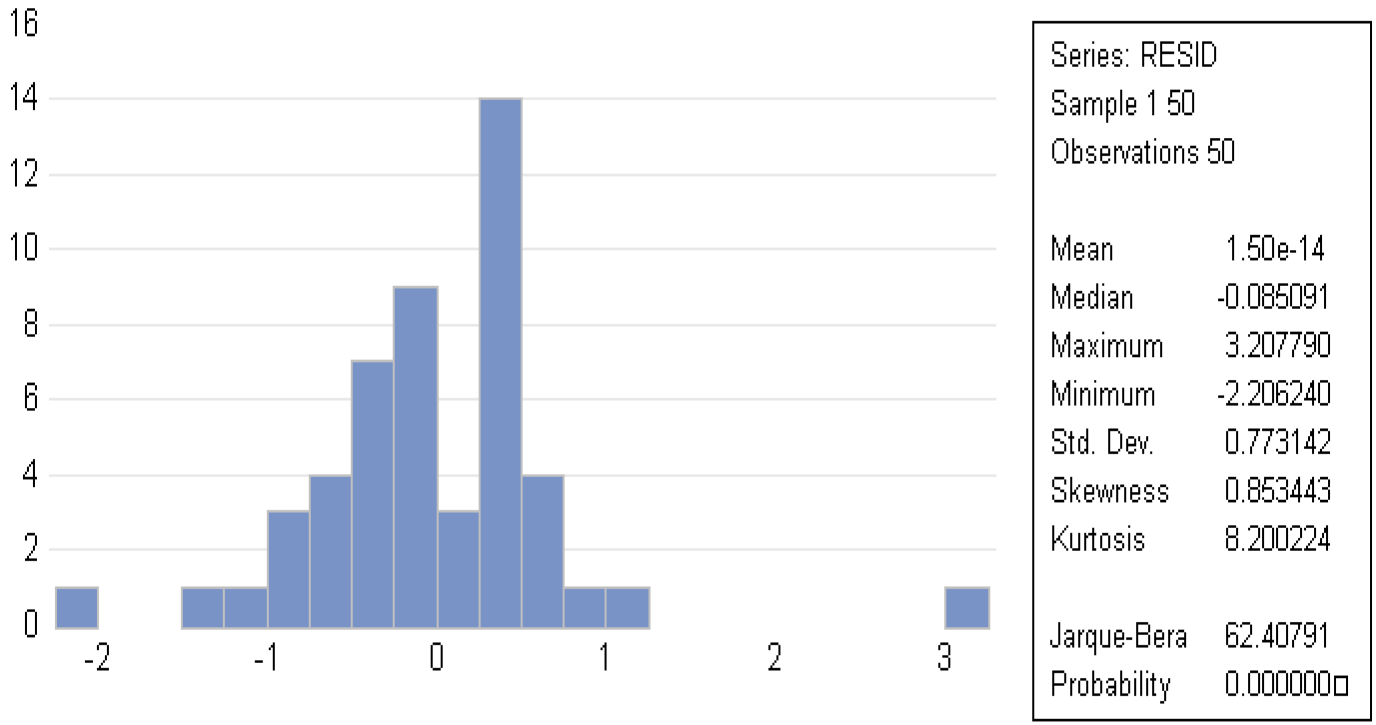


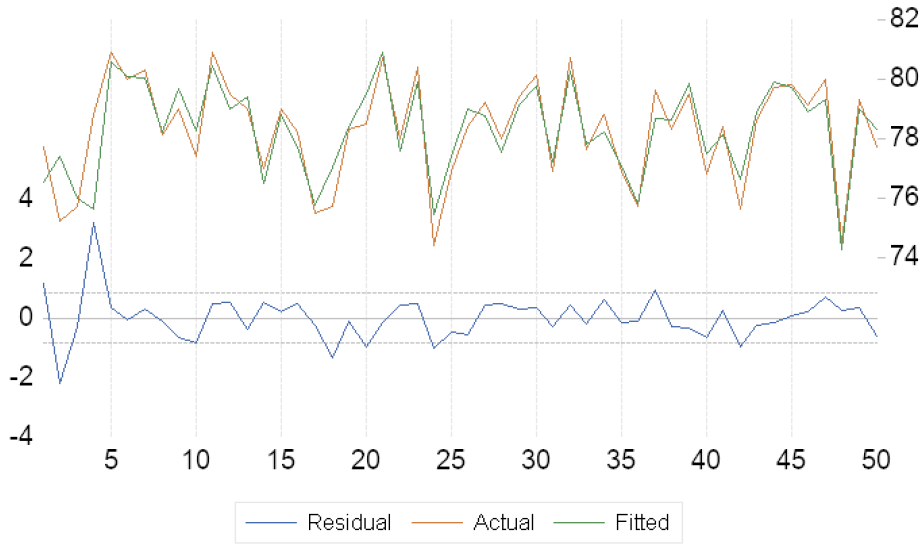
Figure: Eq02 : Jarque-Berra test

F test

- n = 50, k = 10, q = 1, $\widehat{u'_R}\widehat{u_R} = 37.29772$, $\widehat{u'}\widehat{u} = 29.28966$
- so $qF = 10.94 > \chi^2_{0.99} = 6.635$
- we rH_0 and Eq02 cannot be restricted into Eq01
- Inverse quadratic : Life expectancy will grow with sun until a specific value

Residuals

- For example, the model is underestimating the life expectancy of Arkansas ($\widehat{u}_4 = y_4 - \widehat{y}_4 > 0$)
- Although it is one of the poorest states, it is known for its beautiful lakes, rivers, and hot springs which could maybe affect life expectancy



sun

- $\frac{\partial life}{\partial sun} = -0.00074 sun + 0.077$
- 100 sunny days more per year on average will result in an increase of $3 * 10^{-3}$ in life expectancy
- 250 sunny days more per year on average will result in a decrease of -0.11 in life expectancy
- We must be careful with these results

smoking

- $\frac{\partial life}{\partial cig} = -0.24$
- 10 % more smokers in a state will result in a decrease of -2.4 years in its average life expectancy
- It seems logic, this is what we would expect

obesity

- $\frac{\partial life}{\partial obesity} = -0.11$
- 10 % more obese people in a state will result in a decrease of -1.1 years in its average life expectancy
- It seems logic, this is what we would expect

Figure: Queen celebrates 96th birthday in Sandringham