How increase life expectancy ?

Tanguy MALEZIEUX Paul WATFEH

ENAC

April 15, 2019

How increase life expectancy?

In which France departement do we live the longest?

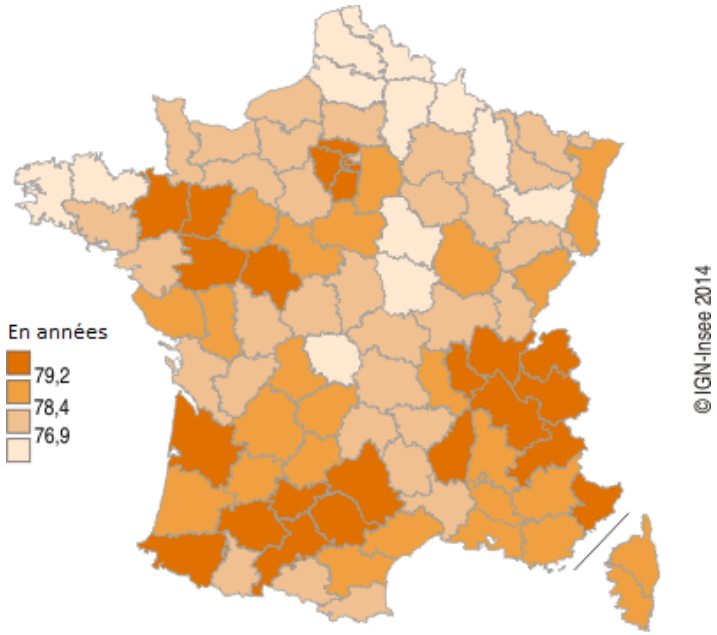


Figure: 2015 life expectancy

Data collecting and hypothesis

- 97 departments
- on the internet: INSEE, government website

Hypothesis 1

The richer the older?

Hypothesis 2

Sun brings life?

Variables	Unit	Sign exp
ESPH	Years from birth	
REVENUE	salary per years	+
POP	Number of inhabitants	?
FEC	total fertility rate	?
ENSUP	graduated people per 1000 inhabitants	+
DROU	people using 2 wheels vehicle per 1000in.	-
TCOM	"" using public transport ""	+
MARC	"" walking ""	+
VOIT	"" car ""	-
MED	doctors per 100,000in.	+
FUMEURS	smokers per 100in.	-
ALCOOL	daily consummer per 100in.	-
POLL	Quality of the air grade from 0 to 20	-
SUN	sun rate in kW per year	+
PLUIE	millimetres per year	-
SPORT	License in main sports per 1000in.	+

Carateristics of the model

- We want to modelize life expectancy of men
- We found this equation

$$\text{EQ01 : y} = \beta_0 + \beta_1 POP * 10^{-4} + \beta_2 FEC + \beta_3 ENSUP + \beta_4 REVENU * 10^{-4} + \beta_5 DROU + \beta_6 TCOM + \beta_7 MARC + \beta_8 AUC + \beta_9 VOIT + \beta_{10} MED + \beta_{11} FUMEURS + \beta_{12} ALCOOL + \beta_{13} POLL + \beta_{14} SUN + \beta_{15} PLUIE + \beta_{16} SPORT$$

First Model

Dependent Variable: ESPH
Method: Least Squares
Date: 04/08/19 Time: 15:21
Sample: 1 97
Included observations: 97

	Coefficient	Std. Error	t-Statistic	Prob.
C	63.60829	2.552291	24.92203	0.0000
POP*(10^-4)	-0.008900	0.002603	-3.419683	0.0010
FEC	0.024446	0.006978	3.503272	0.0008
ENSUP	0.134310	0.044062	3.048218	0.0031
REVENU*(10^-4)	-0.033642	0.213236	-0.157769	0.8750
DROU	0.029817	0.019220	1.551376	0.1248
TCOM	0.020112	0.005915	3.400298	0.0011
MARC	0.003308	0.015672	0.211066	0.8334
AUC	-0.007180	0.025272	-0.284112	0.7771
VOIT	0.017971	0.005065	3.548446	0.0007
MED	0.019097	0.007749	2.464433	0.0159
FUMEURS	-0.055540	0.016348	-3.397385	0.0011
ALCOOL	0.102414	0.046651	2.195310	0.0310
POLL	-0.003227	0.023637	-0.136516	0.8918
SUN	-0.000146	0.000187	-0.781644	0.4367
PLUIE	0.000421	0.000541	0.777889	0.4389
SPORT	0.001279	0.000951	1.344418	0.1826

R-squared	0.717716	Mean dependent var	79.22062
Adjusted R-squared	0.661260	S.D. dependent var	1.219797
S.E. of regression	0.709939	Akaike info criterion	2.310555
Sum squared resid	40.32105	Schwarz criterion	2.761793
Log likelihood	-95.06193	Hannan-Quinn criter.	2.493014
F-statistic	12.71268	Durbin-Watson stat	1.884502
Prob(F-statistic)	0.000000		

Figure: First Model.

Variables	Unit	Sign exp	Sign res
ESPH	Years from birth		
POP	Number of inhabitants	?	-
FEC	total fertility rate	?	+
ENSUP	graduated people per 1000 inhabitants	+	+
DROU	people using 2 wheels vehicle per 1000in.	-	+
TCOM	"" using public transport ""	+	+
MARC	"" walking ""	+	+
VOIT	"" car ""	-	+
MED	doctors per 1000in.	+	+
FUMEURS	smokers per 100in.	-	-
ALCOOL	daily consummer per 100in.	-	+
POLL	Quality of the air grade from 0 to 20	-	-
SUN	sun rate in kW per year	+	+
PLUIE	millimetres per year	-	+
SPORT	License in main sports per 1000in.	+	+

Analysis

- Problems with alcohol
- We add a new variable

EQ02:

$$y=\beta_0 + \beta_1 POP * 10^{-4} + \beta_2 FEC + \beta_3 ENSUP + \beta_4 REVENU * 10^{-4} + \beta_5 DROU + \beta_6 TCOM + \beta_7 MARC + \beta_8 AUC + \beta_9 VOIT + \beta_{10} MED + \beta_{11} FUMEURS + \beta_{12} ALCOOL + \beta_{13} ALCOOL^2 + \beta_{14} POLL + \beta_{15} SUN + \beta_{16} PLUIE + \beta_{17} SPORT$$

Second Model

Dependent Variable: ESPH
Method: Least Squares
Date: 04/08/19 Time: 15:26
Sample: 1 97
Included observations: 97

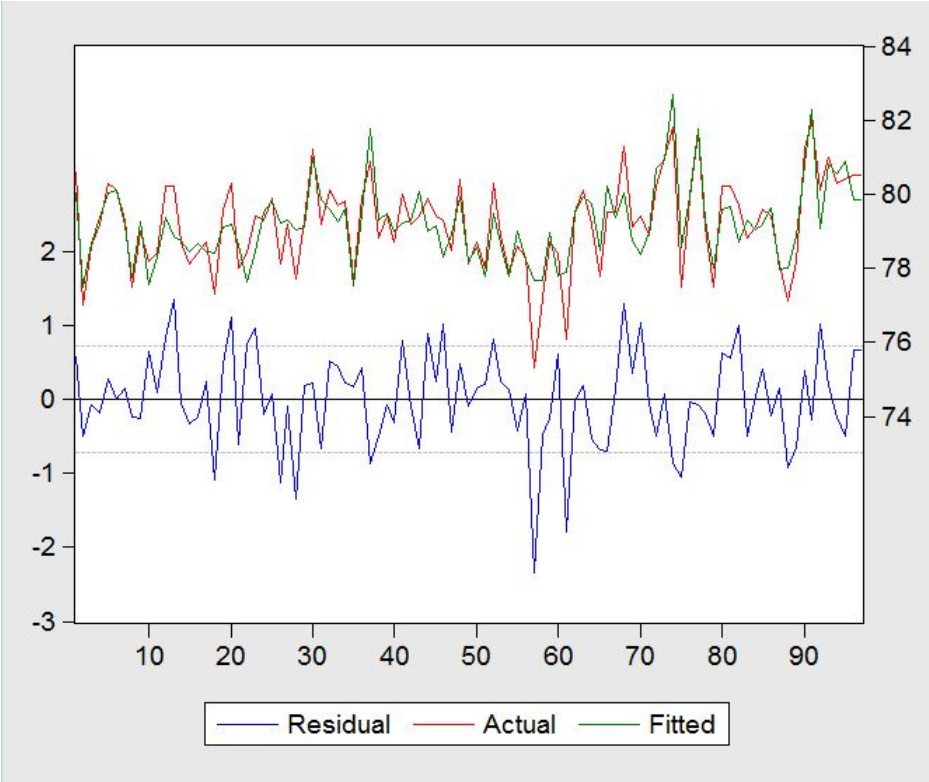
	Coefficient	Std. Error	t-Statistic	Prob.
C	64.82879	4.298911	15.08028	0.0000
POP*(10^-4)	-0.009057	0.002655	-3.412112	0.0010
FEC	0.024298	0.007029	3.456664	0.0009
ENSUP	0.134524	0.044309	3.036056	0.0032
REVENU*(10^-4)	-0.024941	0.215816	-0.115568	0.9083
DROU	0.030656	0.019470	1.574474	0.1194
TCOM	0.020171	0.005950	3.390245	0.0011
MARC	0.003445	0.015763	0.218521	0.8276
AUC	-0.008332	0.025619	-0.325217	0.7459
VOIT	0.018082	0.005102	3.544068	0.0007
MED	0.018893	0.007813	2.418255	0.0179
FUMEURS	-0.056494	0.016658	-3.391480	0.0011
ALCOOL	-0.068102	0.484112	-0.140674	0.8885
ALCOOL^2	0.005876	0.016604	0.353890	0.7244
POLL	-0.005180	0.024399	-0.212296	0.8324
SUN	-0.000127	0.000196	-0.649273	0.5180
PLUIE	0.000430	0.000545	0.790019	0.4319
SPORT	0.001279	0.000956	1.337119	0.1850

R-squared	0.718163	Mean dependent var	79.22062
Adjusted R-squared	0.657515	S.D. dependent var	1.219797
S.E. of regression	0.713852	Akaike info criterion	2.329590
Sum squared resid	40.25723	Schwarz criterion	2.807371
Log likelihood	-94.98510	Hannan-Quinn criter.	2.522781
F-statistic	11.84141	Durbin-Watson stat	1.895603
Prob(F-statistic)	0.000000		

Figure: Second Model.

Residuals

Amplitude for North and Pas de Calais overestimated



Comparison of R-squared

There isn't a too big difference between each model

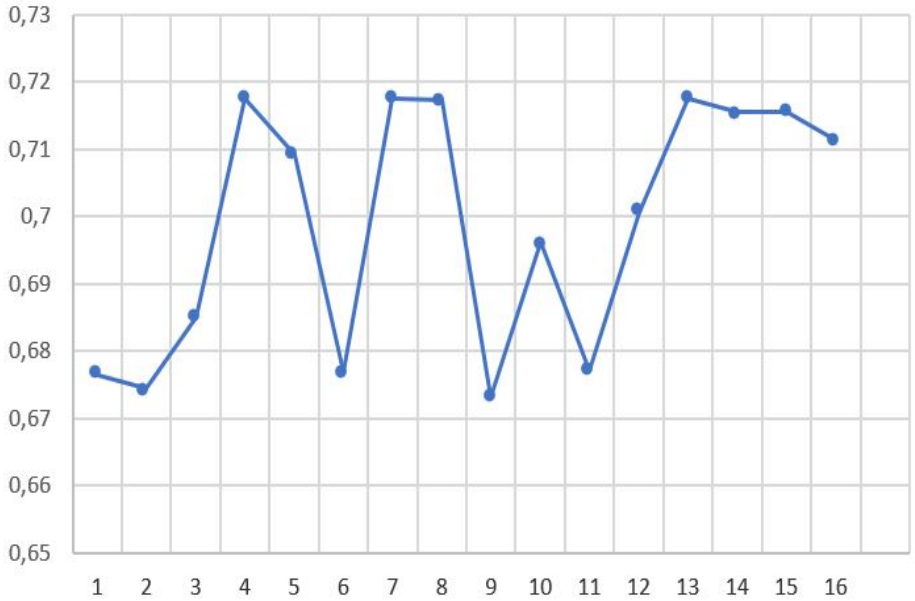


Figure: Residuals.

Model

Dependent Variable: ESPF
Method: Least Squares
Date: 04/14/19 Time: 17:47
Sample: 1 97
Included observations: 97

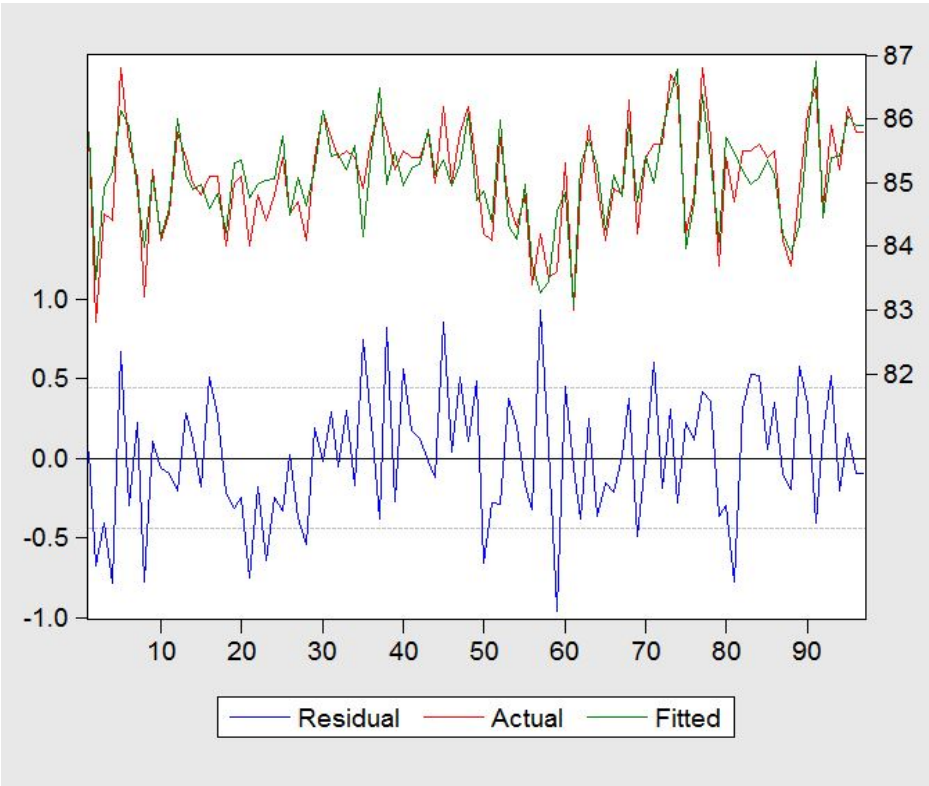
	Coefficient	Std. Error	t-Statistic	Prob.
C	73.16603	1.892495	38.66114	0.0000
POP*(10^-4)	-0.006124	0.001930	-3.173306	0.0021
FEC	0.020597	0.005174	3.980681	0.0002
ENSUP	0.132385	0.032671	4.052017	0.0001
REVENU*(10^-4)	0.246378	0.158112	1.558251	0.1231
DROU	0.035691	0.014251	2.504404	0.0143
TCOM	0.004003	0.004386	0.912831	0.3641
MARC	0.006622	0.011621	0.569868	0.5704
AUC	0.026568	0.018739	1.417785	0.1601
VOIT	0.008110	0.003755	2.159479	0.0338
MED	0.011887	0.005746	2.068814	0.0418
FUMEURS	-0.047809	0.012122	-3.944059	0.0002
ALCOOL	0.051843	0.034591	1.498708	0.1379
POLL	0.012250	0.017526	0.698959	0.4866
SUN	-2.14E-05	0.000139	-0.153868	0.8781
PLUIE	0.000792	0.000401	1.974006	0.0518
SPORT	0.000615	0.000705	0.871434	0.3861

R-squared	0.678838	Mean dependent var	85.08144
Adjusted R-squared	0.614605	S.D. dependent var	0.847955
S.E. of regression	0.526412	Akaike info criterion	1.712365
Sum squared resid	22.16874	Schwarz criterion	2.163603
Log likelihood	-66.04970	Hannan-Quinn criter.	1.894823
F-statistic	10.56845	Durbin-Watson stat	2.133545
Prob(F-statistic)	0.000000		

Figure: Life Expectancy of women.

Residuals

Amplitude for North are overestimated and many britanny departments are underestimated



Analysis

R-squared value

0,72 for men

0,68 for women

Significance

insignificant variables differ from the gender (ALCOOL FUMEURS)

Dependent Variable: ESPH
Method: Least Squares
Date: 04/08/19 Time: 15:21
Sample: 1 97
Included observations: 97

	Coefficient	Std. Error	t-Statistic	Prob.
C	63.60829	2.552291	24.92203	0.0000
POP*(10^-4)	-0.008900	0.002603	-3.419683	0.0010
FEC	0.024446	0.006978	3.503272	0.0008
ENSUP	0.134310	0.044062	3.048218	0.0031
REVENU*(10^-4)	-0.033642	0.213236	-0.157769	0.8750
DROU	0.029817	0.019220	1.551376	0.1248
TCOM	0.020112	0.005915	3.400298	0.0011
MARC	0.003308	0.015672	0.211066	0.8334
AUC	-0.007180	0.025272	-0.284112	0.7771
VOIT	0.017971	0.005065	3.548446	0.0007
MED	0.019097	0.007749	2.464433	0.0159
FUMEURS	-0.055540	0.016348	-3.397385	0.0011
ALCOOL	0.102414	0.046651	2.195310	0.0310
POLL	-0.003227	0.023637	-0.136516	0.8918
SUN	-0.000146	0.000187	-0.781644	0.4367
PLUIE	0.000421	0.000541	0.777889	0.4389
SPORT	0.001279	0.000951	1.344418	0.1826

R-squared	0.717716	Mean dependent var	79.22062
Adjusted R-squared	0.661260	S.D. dependent var	1.219797
S.E. of regression	0.709939	Akaike info criterion	2.310555
Sum squared resid	40.32105	Schwarz criterion	2.761793
Log likelihood	-95.06193	Hannan-Quinn criter.	2.493014
F-statistic	12.71268	Durbin-Watson stat	1.884502
Prob(F-statistic)	0.000000		

Figure: ESPH

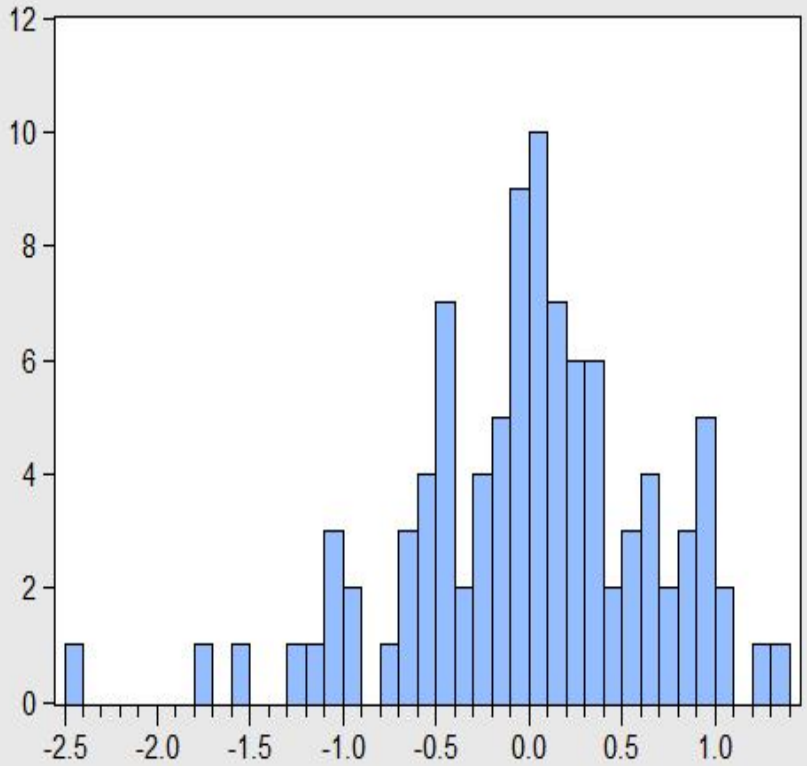
Dependent Variable: ESPF
Method: Least Squares
Date: 04/14/19 Time: 17:47
Sample: 1 97
Included observations: 97

	Coefficient	Std. Error	t-Statistic	Prob.
C	73.16603	1.892495	38.66114	0.0000
POP*(10^-4)	-0.006124	0.001930	-3.173306	0.0021
FEC	0.020597	0.005174	3.980681	0.0002
ENSUP	0.132385	0.032671	4.052017	0.0001
REVENU*(10^-4)	0.246378	0.158112	1.558251	0.1231
DROU	0.035691	0.014251	2.504404	0.0143
TCOM	0.004003	0.004386	0.912831	0.3641
MARC	0.006622	0.011621	0.569868	0.5704
AUC	0.026568	0.018739	1.417785	0.1601
VOIT	0.008110	0.003755	2.159479	0.0338
MED	0.011887	0.005746	2.068814	0.0418
FUMEURS	-0.047809	0.012122	-3.944059	0.0002
ALCOOL	0.051843	0.034591	1.498708	0.1379
POLL	0.012250	0.017526	0.698959	0.4866
SUN	-2.14E-05	0.000139	-0.153868	0.8781
PLUIE	0.000792	0.000401	1.974006	0.0518
SPORT	0.000615	0.000705	0.871434	0.3861

R-squared	0.678838	Mean dependent var	85.08144
Adjusted R-squared	0.614605	S.D. dependent var	0.847955
S.E. of regression	0.526412	Akaike info criterion	1.712365
Sum squared resid	22.16874	Schwarz criterion	2.163603
Log likelihood	-66.04970	Hannan-Quinn criter.	1.894823
F-statistic	10.56845	Durbin-Watson stat	2.133545
Prob(F-statistic)	0.000000		

Figure: ESPF

Normality of the error



Series: RESID	
Sample 1 97	
Observations 97	
Mean	-1.67e-14
Median	0.046915
Maximum	1.320301
Minimum	-2.402305
Std. Dev.	0.655363
Skewness	-0.671224
Kurtosis	4.216591
Jarque-Bera	13.26581
Probability	0.001316

Homoscedasticity

Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	1.417124	Prob. F(17,79)	0.1511
Obs*R-squared	22.66769	Prob. Chi-Square(17)	0.1604
Scaled explained SS	23.32526	Prob. Chi-Square(17)	0.1390

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 04/11/19 Time: 08:31
Sample: 1 97
Included observations: 97

	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.246164	4.270381	-0.057645	0.9542
POP*(10^-4)	0.006757	0.002637	2.562656	0.0123
FEC	-0.001484	0.006983	-0.212551	0.8322
ENSUP	-0.006566	0.044015	-0.149168	0.8818
REVENU*(10^-4)	0.065255	0.214384	0.304383	0.7616
DROU	-0.001461	0.019341	-0.075527	0.9400
TCOM	-0.011705	0.005910	-1.980502	0.0511
MARC	0.008324	0.015658	0.531604	0.5965
AUC	0.009705	0.025449	0.381333	0.7040
VOIT	-0.008722	0.005068	-1.720929	0.0892
MED	-0.014732	0.007761	-1.898213	0.0613
FUMEURS	-0.009264	0.016547	-0.559858	0.5772
ALCOOL	0.731313	0.480899	1.520721	0.1323
ALCOOL^2	-0.022226	0.016493	-1.347597	0.1816
POLL	0.012941	0.024237	0.533914	0.5949
SUN	3.46E-05	0.000195	0.177889	0.8593
PLUIE	-0.000231	0.000541	-0.427713	0.6700
SPORT	-0.001386	0.000950	-1.459041	0.1485

R-squared	0.233688	Mean dependent var	0.415023
Adjusted R-squared	0.068785	S.D. dependent var	0.734838
S.E. of regression	0.709115	Akaike info criterion	2.316272
Sum squared resid	39.72467	Schwarz criterion	2.794054
Log likelihood	-94.33921	Hannan-Quinn criter.	2.509464
F-statistic	1.417124	Durbin-Watson stat	2.551835
Prob(F-statistic)	0.151134		

Conclusion