School of Engineering

An investigation on asthma risk factors

Empirical Projet - Econometrics I

Rémi Christien
Samy Doreau

Introduction

Why this study?

Asthma = still a mystery for medicine

Known risk factors :

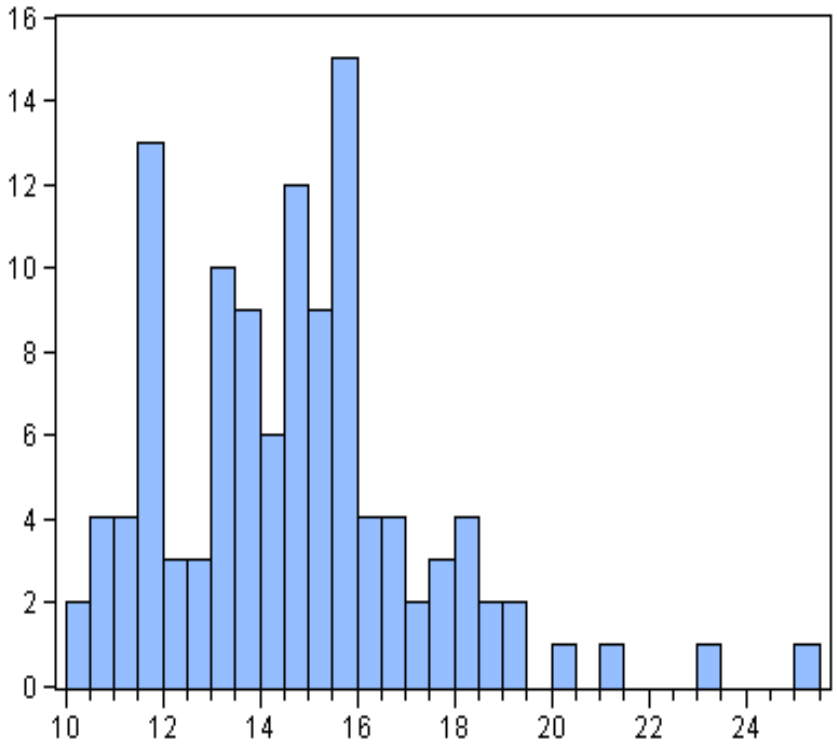
- allergic condition
- being overweight, or a smoker
- exposure to exhaust fumes
- ...

Outline

1. Presentation of parameters
2. Two models
3. Conclusion

Our first parameter

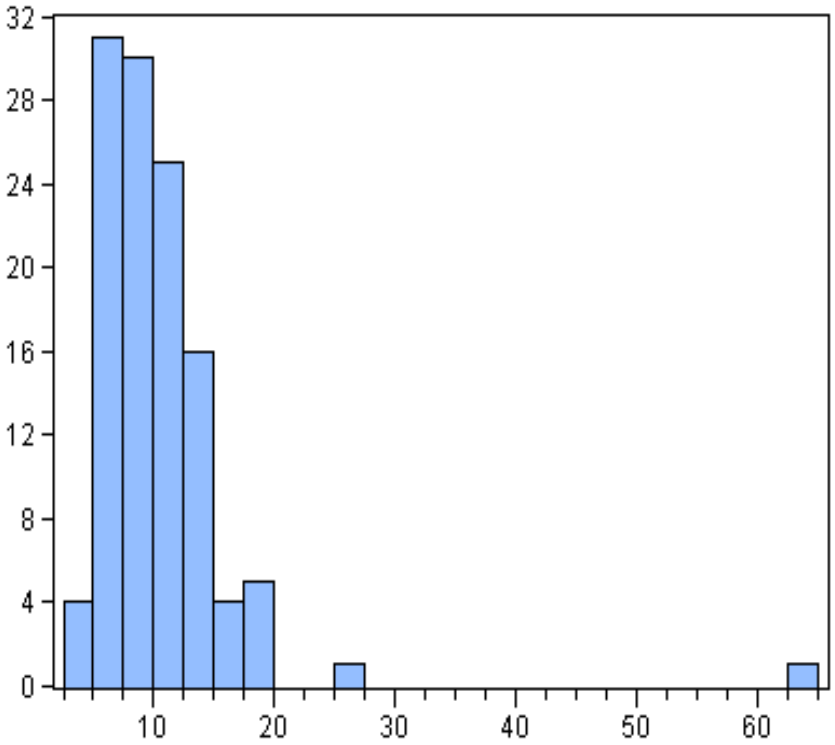
Asthma prevalence



Series: PREV	
Sample 1 120	
Observations 115	
Mean	14.62261
Median	14.60000
Maximum	25.00000
Minimum	10.10000
Std. Dev.	2.616012
Skewness	0.951195
Kurtosis	4.863102
Jarque-Bera	33.97404
Probability	0.000000

Second parameter

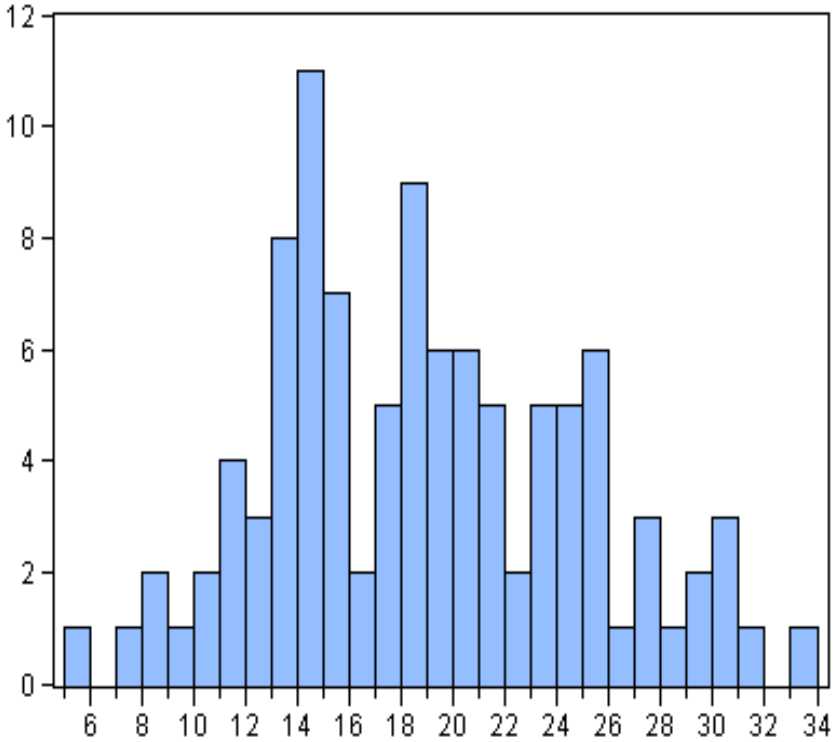
Asthma related hospitalizations



Series: HOSP	
Sample 1 120	
Observations 117	
Mean	10.52479
Median	9.500000
Maximum	63.90000
Minimum	3.200000
Std. Dev.	6.236460
Skewness	5.580244
Kurtosis	47.23270
Jarque-Bera	10145.30
Probability	0.000000

Third parameter

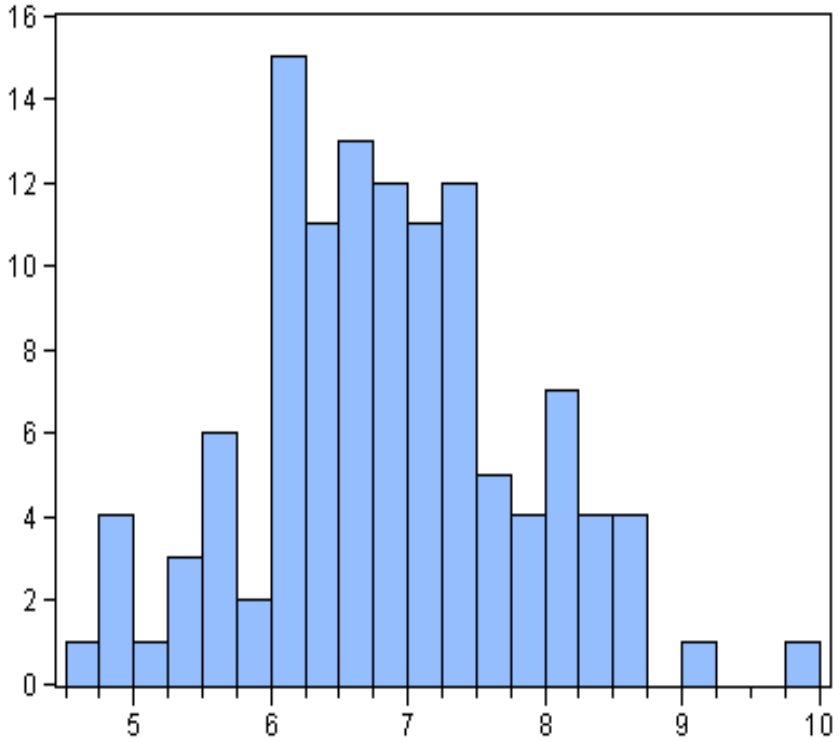
Smokers rate



Series: SMOK	
Sample 1 120	
Observations 103	
Mean	18.38835
Median	18.00000
Maximum	33.00000
Minimum	5.000000
Std. Dev.	5.957748
Skewness	0.268250
Kurtosis	2.491979
Jarque-Bera	2.342895
Probability	0.309918

Fourth parameter

Low birth weight rate

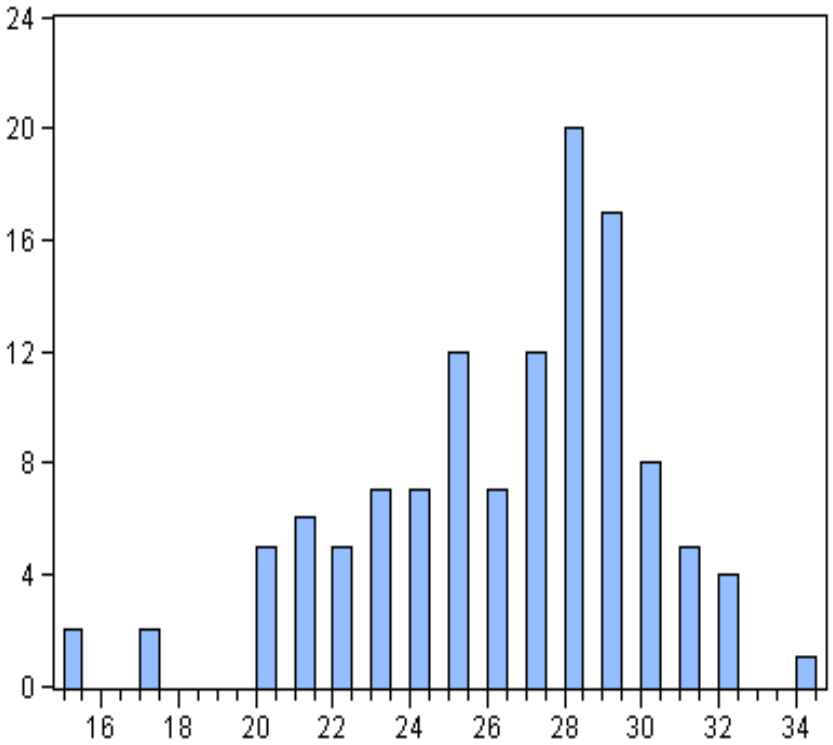


Series: BIRTH	
Sample 1 120	
Observations 117	
Mean	6.830769
Median	6.800000
Maximum	9.800000
Minimum	4.600000
Std. Dev.	0.972507
Skewness	0.187258
Kurtosis	3.033425
Jarque-Bera	0.689228
Probability	0.708494

Figure: $\Longrightarrow$ Normal distribution after a Jarque Bera test

Fifth parameter

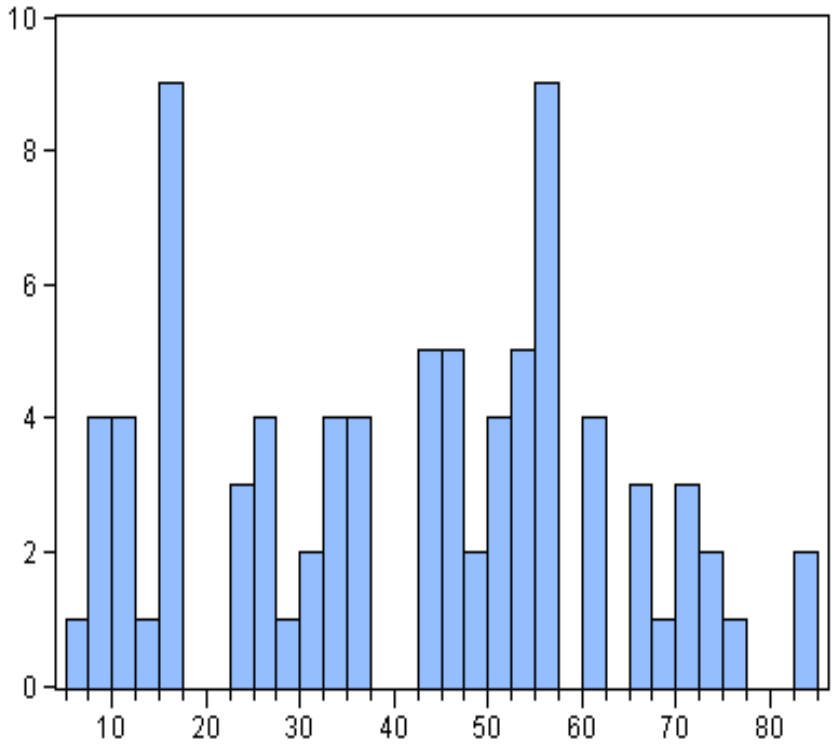
Obesity rate



Series: OBES	
Sample 1 120	
Observations 120	
Mean	26.20833
Median	27.00000
Maximum	34.00000
Minimum	15.00000
Std. Dev.	3.687125
Skewness	-0.739922
Kurtosis	3.393248
Jarque-Bera	11.72291
Probability	0.002847

Sixth parameter

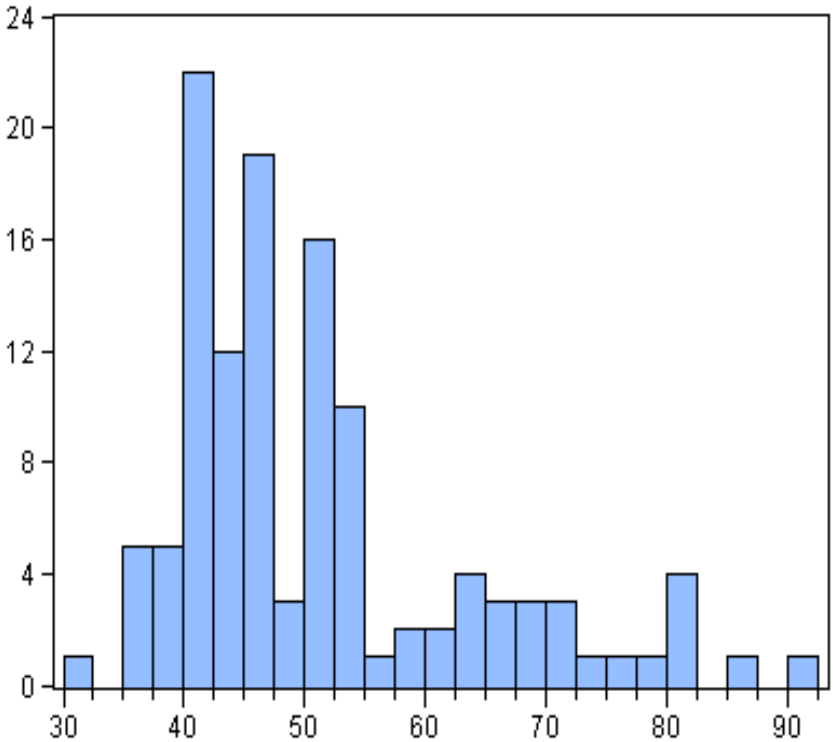
Air quality index



Series: AIR	
Sample 1 120	
Observations 83	
Mean	41.17229
Median	44.00000
Maximum	83.00000
Minimum	5.000000
Std. Dev.	20.82711
Skewness	-0.034853
Kurtosis	1.956595
Jarque-Bera	3.781873
Probability	0.150930

Seventh parameter

Household income



Series: INC	
Sample 1 120	
Observations 120	
Mean	50.67500
Median	46.00000
Maximum	90.00000
Minimum	32.00000
Std. Dev.	12.25442
Skewness	1.270706
Kurtosis	4.020194
Jarque-Bera	37.49788
Probability	0.000000

Correlation table

	SMOK	PREV	OBES	INC	HOSP	BIRTH	AIR
SMOK	1.000000	0.445680	0.589851	-0.492477	0.230164	0.401208	-0.352120
PREV	0.445680	1.000000	0.486431	-0.412364	0.169789	0.275636	-0.086186
OBES	0.589851	0.486431	1.000000	-0.583436	0.297132	0.247668	-0.309312
INC	-0.492477	-0.412364	-0.583436	1.000000	0.032746	0.063253	0.043453
HOSP	0.230164	0.169789	0.297132	0.032746	1.000000	0.557083	-0.015126
BIRTH	0.401208	0.275636	0.247668	0.063253	0.557083	1.000000	-0.077673
AIR	-0.352120	-0.086186	-0.309312	0.043453	-0.015126	-0.077673	1.000000

Asthma prevalence

Equation

$$\widehat{PREV_i} = \widehat{\alpha_0} + \widehat{\alpha_1} \times SMOK + \widehat{\alpha_2} \times BIRTH + \widehat{\alpha_3} \times OBES + \widehat{\alpha_4} \times AIR + \widehat{\alpha_5} \times INCOME$$

1st Equation

Dependent Variable: PREV
Method: Least Squares
Date: 05/08/12 Time: 22:34
Sample (adjusted): 1 118
Included observations: 68 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
C	7.461172	3.442066	2.167643	0.0340
OBES	0.155602	0.089607	1.736486	0.0874
BIRTH	0.486835	0.361088	1.348246	0.1825
AIR	0.007829	0.013439	0.582571	0.5623
INC	-0.034834	0.025936	-1.343099	0.1841
SMOK	0.065219	0.067266	0.969559	0.3360

R-squared	0.316777	Mean dependent var	14.31324
Adjusted R-squared	0.261678	S.D. dependent var	2.427695
S.E. of regression	2.086013	Akaike info criterion	4.392484
Sum squared resid	269.7900	Schwarz criterion	4.588323
Log likelihood	-143.3444	Hannan-Quinn criter.	4.470081
F-statistic	5.749260	Durbin-Watson stat	1.729383
Prob(F-statistic)	0.000202		

Wald test

Wald Test:
Equation: EQ1PREV

Test Statistic	Value	df	Probability
F-statistic	0.511135	(2, 62)	0.6023
Chi-square	1.022270	2	0.5998

Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(3)	0.007829	0.013439
C(6)	0.065219	0.067266

Restrictions are linear in coefficients.

Figure: Can we take out AIR and SMOKE ? ... YES

2nd Equation

Dependent Variable: PREV
Method: Least Squares
Date: 05/09/12 Time: 21:07
Sample: 1 120
Included observations: 112

	Coefficient	Std. Error	t-Statistic	Prob.
C	9.272800	2.977627	3.114158	0.0024
BIRTH	0.566665	0.264752	2.140364	0.0346
INC	-0.046111	0.023040	-2.001330	0.0479
OBES	0.148331	0.079885	1.856805	0.0661

R-squared	0.187243	Mean dependent var	14.66518
Adjusted R-squared	0.164666	S.D. dependent var	2.631161
S.E. of regression	2.404792	Akaike info criterion	4.627865
Sum squared resid	624.5668	Schwarz criterion	4.724954
Log likelihood	-255.1604	Hannan-Quinn criter.	4.667257
F-statistic	8.293666	Durbin-Watson stat	1.891664
Prob(F-statistic)	0.000052		

Figure: PREV = f(INC, LBW, OBES)

Heteroskedasticity

Heteroskedasticity Test: White

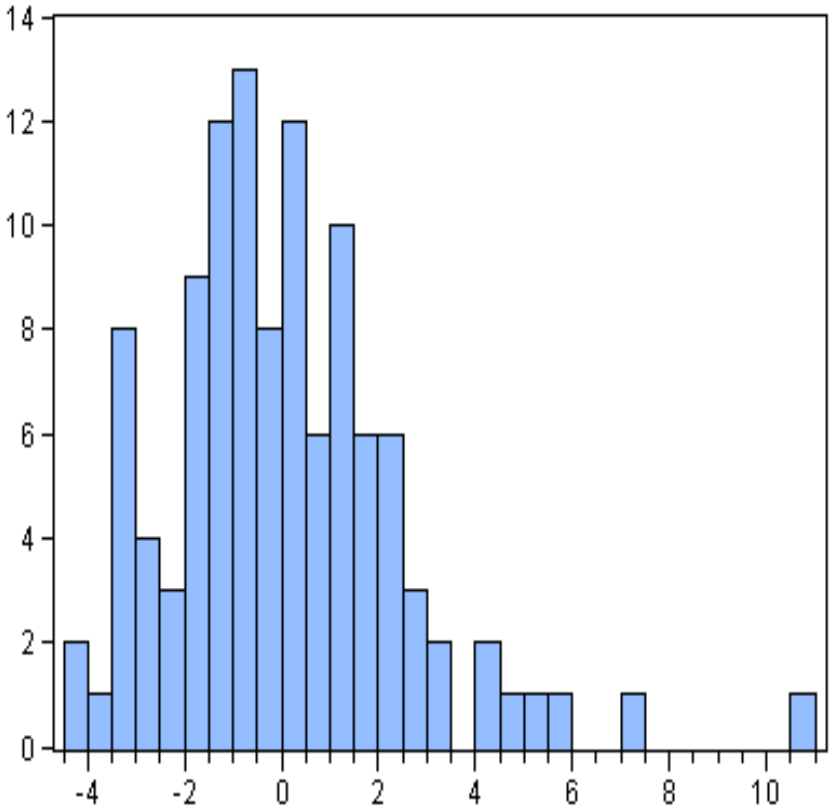
F-statistic	0.726219	Prob. F(9,102)	0.6839
Obs*R-squared	6.744571	Prob. Chi-Square(9)	0.6637
Scaled explained SS	16.71542	Prob. Chi-Square(9)	0.0534

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 05/09/12 Time: 21:56
Sample: 1 120
Included observations: 112

	Coefficient	Std. Error	t-Statistic	Prob.
C	-196.6879	188.0446	-1.045964	0.2981
BIRTH	30.64673	25.34216	1.209318	0.2293
BIRTH^2	-1.250955	1.265062	-0.988849	0.3251
BIRTH*INC	-0.048483	0.227385	-0.213221	0.8316
BIRTH*OBES	-0.390047	0.592026	-0.658835	0.5115
INC	1.397195	1.888384	0.739889	0.4611
INC^2	-0.011684	0.009617	-1.214931	0.2272
INC*OBES	0.009237	0.053537	0.172540	0.8634
OBES	5.063162	9.195917	0.550588	0.5831
OBES^2	-0.063501	0.132185	-0.480396	0.6320

R-squared	0.060219	Mean dependent var	5.576489
Adjusted R-squared	-0.022702	S.D. dependent var	12.93299
S.E. of regression	13.07898	Akaike info criterion	8.064935
Sum squared resid	17448.08	Schwarz criterion	8.307658
Log likelihood	-441.6363	Hannan-Quinn criter.	8.163415
F-statistic	0.726219	Durbin-Watson stat	2.213535
Prob(F-statistic)	0.683912		

Residus



Series: RESID	
Sample 1 120	
Observations 112	
Mean	1.99e-16
Median	-0.251634
Maximum	10.82448
Minimum	-4.312179
Std. Dev.	2.372072
Skewness	1.209670
Kurtosis	6.330666
Jarque-Bera	79.08386
Probability	0.000000

Asthma related hospitalizations

Equation

$$\widehat{HOSP_i} = \widehat{\alpha_0} + \widehat{\alpha_1} \times SMOK + \widehat{\alpha_2} \times BIRTH + \widehat{\alpha_3} \times OBES + \widehat{\alpha_4} \times AIR + \widehat{\alpha_5} \times INCOME$$

1st Equation

Dependent Variable: HOSP
Method: Least Squares
Date: 05/08/12 Time: 22:36
Sample (adjusted): 1 118
Included observations: 72 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
C	-12.36009	9.100872	-1.358121	0.1790
AIR	-0.013178	0.036420	-0.361827	0.7186
BIRTH	5.658831	0.795920	7.109802	0.0000
INC	-0.139402	0.063867	-2.182692	0.0326
OBES	-0.042448	0.221513	-0.191627	0.8486
SMOK	-0.343948	0.177495	-1.937791	0.0569

R-squared	0.446354	Mean dependent var	11.43056
Adjusted R-squared	0.404411	S.D. dependent var	7.425150
S.E. of regression	5.730317	Akaike info criterion	6.409074
Sum squared resid	2167.211	Schwarz criterion	6.598796
Log likelihood	-224.7267	Hannan-Quinn criter.	6.484603
F-statistic	10.64194	Durbin-Watson stat	0.630866
Prob(F-statistic)	0.000000		

Wald test

Wald Test:
Equation: EQ1HOSP

Test Statistic	Value	df	Probability
F-statistic	0.072394	(2, 66)	0.9302
Chi-square	0.144788	2	0.9302

Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(2)	-0.013178	0.036420
C(5)	-0.042448	0.221513

Restrictions are linear in coefficients.

Figure: $\Longrightarrow$ index and obesity rate OUT !

2nd Equation

Dependent Variable: HOSP
Method: Least Squares
Date: 05/08/12 Time: 23:13
Sample (adjusted): 1 119
Included observations: 102 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
C	-10.44011	5.320508	-1.962239	0.0526
BIRTH	4.146497	0.654647	6.333944	0.0000
INC	-0.081832	0.051296	-1.595284	0.1139
SMOK	-0.170761	0.115222	-1.482022	0.1415

R-squared	0.301090	Mean dependent var	10.96078
Adjusted R-squared	0.279695	S.D. dependent var	6.536162
S.E. of regression	5.547292	Akaike info criterion	6.302923
Sum squared resid	3015.700	Schwarz criterion	6.405863
Log likelihood	-317.4491	Hannan-Quinn criter.	6.344607
F-statistic	14.07279	Durbin-Watson stat	1.985920
Prob(F-statistic)	0.000000		

Figure: HOSP : f(LBW, SMOKE, INCOME)

3rd Equation

Dependent Variable: HOSP
Method: Least Squares
Date: 05/08/12 Time: 22:39
Sample: 1 120
Included observations: 116

	Coefficient	Std. Error	t-Statistic	Prob.
C	-12.18635	3.585318	-3.398959	0.0009
BIRTH	3.338878	0.520927	6.409494	0.0000

R-squared	0.264903	Mean dependent var	10.56983
Adjusted R-squared	0.258455	S.D. dependent var	6.244374
S.E. of regression	5.377219	Akaike info criterion	6.219311
Sum squared resid	3296.251	Schwarz criterion	6.266786
Log likelihood	-358.7200	Hannan-Quinn criter.	6.238583
F-statistic	41.08161	Durbin-Watson stat	1.986995
Prob(F-statistic)	0.000000		

Figure: HOSP = f(BIRTH)

Heteroskedasticity

Heteroskedasticity Test: White

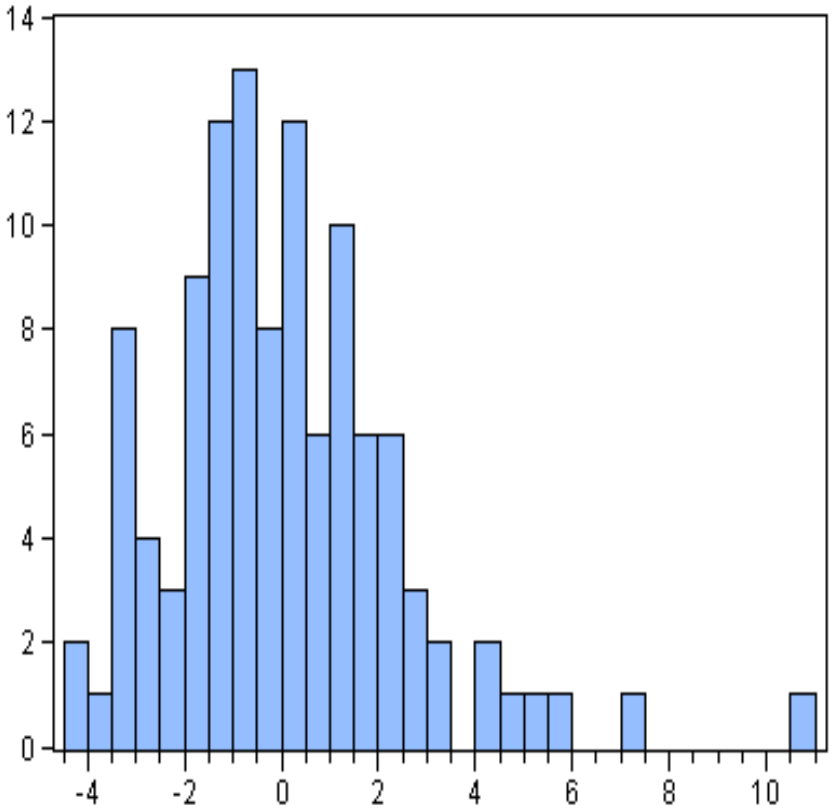
F-statistic	40.71512	Prob. F(2,113)	0.0000
Obs*R-squared	48.58251	Prob. Chi-Square(2)	0.0000
Scaled explained SS	886.0134	Prob. Chi-Square(2)	0.0000

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 05/09/12 Time: 21:57
Sample: 1 120
Included observations: 116

	Coefficient	Std. Error	t-Statistic	Prob.
C	3068.699	452.8825	6.775927	0.0000
BIRTH	-958.6904	131.7456	-7.276829	0.0000
BIRTH^2	73.75345	9.491786	7.770240	0.0000

R-squared	0.418815	Mean dependent var	28.41596
Adjusted R-squared	0.408528	S.D. dependent var	175.3843
S.E. of regression	134.8831	Akaike info criterion	12.67222
Sum squared resid	2055861.	Schwarz criterion	12.74343
Log likelihood	-731.9886	Hannan-Quinn criter.	12.70113
F-statistic	40.71512	Durbin-Watson stat	2.125089
Prob(F-statistic)	0.000000		

and to finish with ...Residus !



Series: RESID	
Sample 1 120	
Observations 112	
Mean	1.99e-16
Median	-0.251634
Maximum	10.82448
Minimum	-4.312179
Std. Dev.	2.372072
Skewness	1.209670
Kurtosis	6.330666
Jarque-Bera	79.08386
Probability	0.000000

In a nutshell ...

- Two models
- Linearity, Strict Endogeneity, No-multicolinearity, Unconditionnal homoskedasticity, and normality of errors
- Normality of errors and heteroskedasticity ... limits of our model?
- LBW, Household income, Obesity rate stand out as influent parameters

Thank you for your attention