

----- *Questions* -----

Is there a contextual effect influencing the residential household energy consumption in France?

How much of the variation of the household energy consumption is due to the differences between the 81 french administrative division so-called “départements”?

----- *Abstract* -----

Multilevel regression modelling introduces simultaneously two levels of aggregation in the model, and offers the possibility of extracting contextual effects from total variation of the household energy consumption. The remaining variation is explained by the model using relevant level-1 (households) explanatory variables.

Using this Shiny App, one can predict the yearly household energy consumption, playing interactively with numerous level-1 predictors, based on a random-intercept multilevel model that can be transposed with the following equation:

$$Y_{ij} = \gamma_{00} + \sum_k \beta_{k0} \cdot X_{kij} + \sum_l \beta_{0l} \cdot Z_{lj} + u_{0j} + e_{ij}$$

Where:

$Y_{ij}$  is the annual energy consumption of household  $i$  in geographical division  $j$  (DEP).

$X_{ij}$  is the matrix of level-1 explanatory variables.

$Z_j$  is the matrix of level-2 explanatory variables.

Other parameters in the equation need to be estimated.

For fixed effects,  $\beta_{00}$  is the intercept for fixed effects,  $\beta_{k0}$  is the slope for level-1 explanatory variables, and  $\beta_{0l}$  is the slope for level-2 explanatory variables.

For random effects,  $e_{ij}$  is the residual term at level-1 (households) and  $u_{0j}$  is the residual term at level-2 (DEP). The variance of the residual error  $u_{0j}$  is the variance of the intercepts between DEPs.

----- *Multilevel Model with R* -----

Using the nlme package, the two-level random-intercept model can be fitted with the following command line:

```
model <- lme(LOGCONSTOT ~      LOGHDDDEP+
                              LOGREV+
                              AREA3G+
                              INSULHOUS+
                              YEARCONST+
                              ROOMNBR+
                              HEATSYST+
                              HEATSOURCE+
                              RURAL+
                              HEATTEMP+
                              ECS+
                              UNOCCWEEK+
                              PCS+
                              NBRPERS,
              random = ~ 1 | DEP,
              data = phebus)
```

----- *Dictionary of codes* -----

**LOGCONSTOT** is the variable response indicating the logarithm of the energy consumption in 2012, given by each household interviewed during Phebus survey. This value is taking in account the energy used for space-heating, space cooling, and electrical appliance (source of data INSEE).

**LOGREVDEP** is a numerical variable indicating the logarithm of the average per capita disposable household income, per DEP in 2012 (source of data INSEE).

**LOGHDDDEP** is a numerical variable indicating logarithm of the heating degree days in 2012 for each department (source of data ADEME).

**LOGREV** is a numerical variable indicating the gross household income disposed in 2012, with a log transformation to reduce a right skewed distribution.

**AREA3G** is a categorical variable indicating the area of the housing, divided in three groups:  
0 - 40 m<sup>2</sup> ; 40 m<sup>2</sup> - 100 m<sup>2</sup> ; 100m<sup>2</sup> and above.

**INSULHOUS** is a binary indicator measuring whether the housing is insulated (=1) or adjoining to another housing unit (=0).

**YEARCONST** is a numerical variable indicating the year when the house was built.

**ROOMNBR** is a numerical variable indicating the number of bedrooms in the housing.

**HEATSYST** is a categorical variable indicating whether the space heating system is dedicated only for heating the housing, for heating a cluster of housing (collective space heating system), or a mixed system (individual and collective).

**HEATSOURCE** indicates the type of energy used for the space heating system. Three categories are defined: electricity ; gas ; other.

**RURAL** is a binary indicator showing if the housing is located in a rural area (i.e. less than 2000 hab). 1= Yes, 0 = No.

**HEATTEMP** is another binary variable indicating the heating temperature selected by the households to heat their housing is above 21°C (included). 1= Yes, 0= No.

**ECS** is a categorical variable indicating how is produced hot water. Three categories are defined: Electricity, boiler (using gas, fuel, or wood as energy source), and others.

**UNOCCWEEK** is a binary variable indicating whether the housing is unoccupied less than four hours during weekdays. 1= Yes, 0= No.

**PCS** is a categorical variable indicating the household employment status. Three categories are defined: Executive status, Middle-level status, and other status.

**NBRPERS** is a numerical variable indicating the number of persons living in housing.