PREDICTIVE MODELS FOR THE NUMBER OF STUDENTS AT YOUTH SCHOOL

Pedro Martins, Luísa Canto e Castro, Nuno Rodrigues, Joana Duarte e Joaquim Santos
www.dgeec.mec.pt

DGEEC
DIREÇÃO-GERAL DE ESTATÍSTICAS DA EDUCAÇÃO E CIÊNCIA

OBJECTIVE

To establish consistent models to predict the total number of students in the school system, that includes the Portuguese demographic regression and the enlargement of the compulsory education.

MAIN ASSUMPTIONS

Analysis limited to regular and artistic courses.
Graduation rates and enrolment age structure seen as constants.

Prediction Model

The prediction model works for students ranging from the 1st cycle of compulsory education (Primary education) to the 12th grade of the Upper Secondary education (ISCED 3), and provides us with the total number of students enrolled without a distinction between students from public schools and students from private schools.
The prediction models works iteratively in the following way

$$\hat{X}_{i,s}^{(a)} = \qquad \tilde{R}_{X,1}^{(a)} = X_{i,0}^{(a)}\left(1 + \beta_{X,i}^{(a)} \times c_0^{(a)}\right) \Big/ P_{-(a-6)}^{(6)}$$

for $a = 6, 7, \dots, 14$; $s = 1, 2, \dots, k$ where k is the number of years to be predicted.

$$\hat{X}_{i,1}^{(a)} = \tilde{R}_{X,1}^{(a)} \times P_{-(a-6+1)}^{(6)} \text{ for } a = 15, 16, 17 \ e \ 18$$

where

$$\tilde{R}_{X,1}^{(a)} = X_{i,0}^{(a)}\left(1 + \beta_{X,i}^{(a)} \times c_0^{(a)}\right) \Big/ P_{-(a-6)}^{(6)}$$

$$\hat{X}_{i,s}^{(a)} = \tilde{R}_{X,s}^{(a)} \times P_{-(a-6+1)}^{(6)} \text{ for } a = 15, 16, 17 \ e \ 18$$

Where

$$\tilde{R}_{X,s}^{(a)} = \hat{X}_{i,s-1}^{(a)}\left(1 + \beta_{X,i}^{(a)} \times c_{s-1}^{(a)}\right) \Big/ P_{-(a-6)}^{(6)}$$

$P_t^{(6)}$ - total of children under 6 years old.
$X_{i,0}^{(a)}$ - number of young people with a years old to attend the i-th grade of the training offer X in year 0.
$\hat{X}_{i,t}^{(a)}$ - prediction for the number of young people aged a who will attend the ith grade educational supply X in year t.
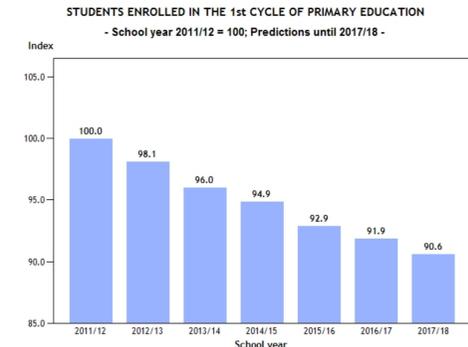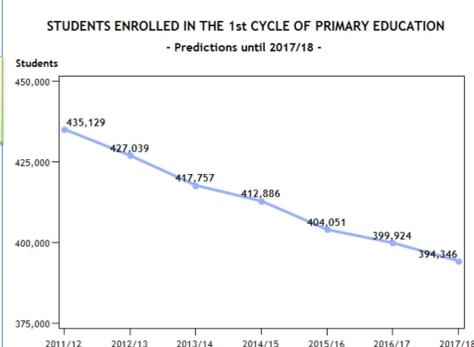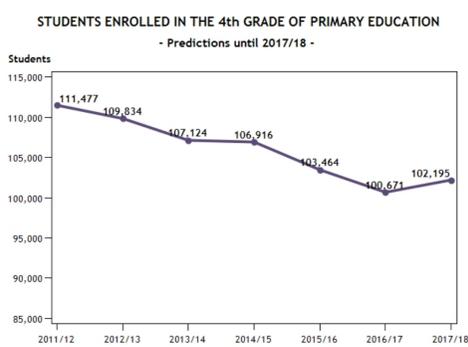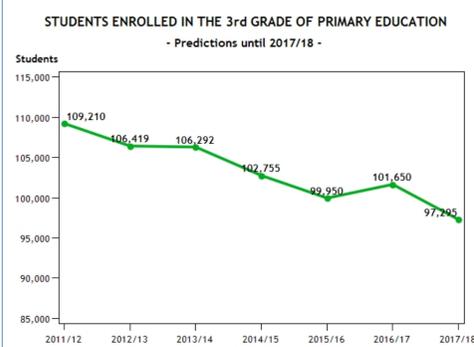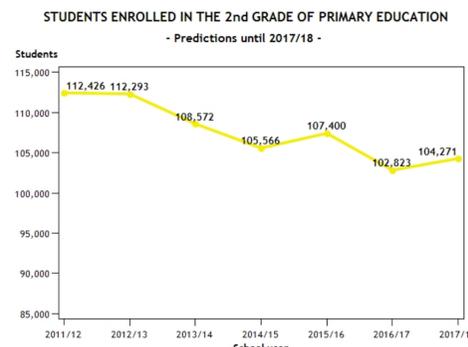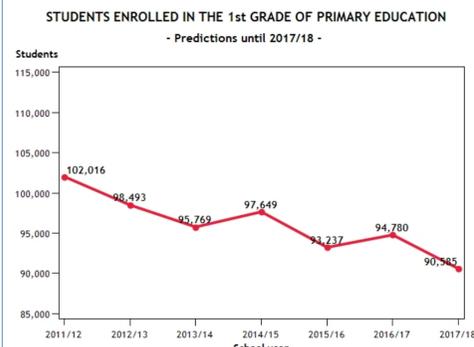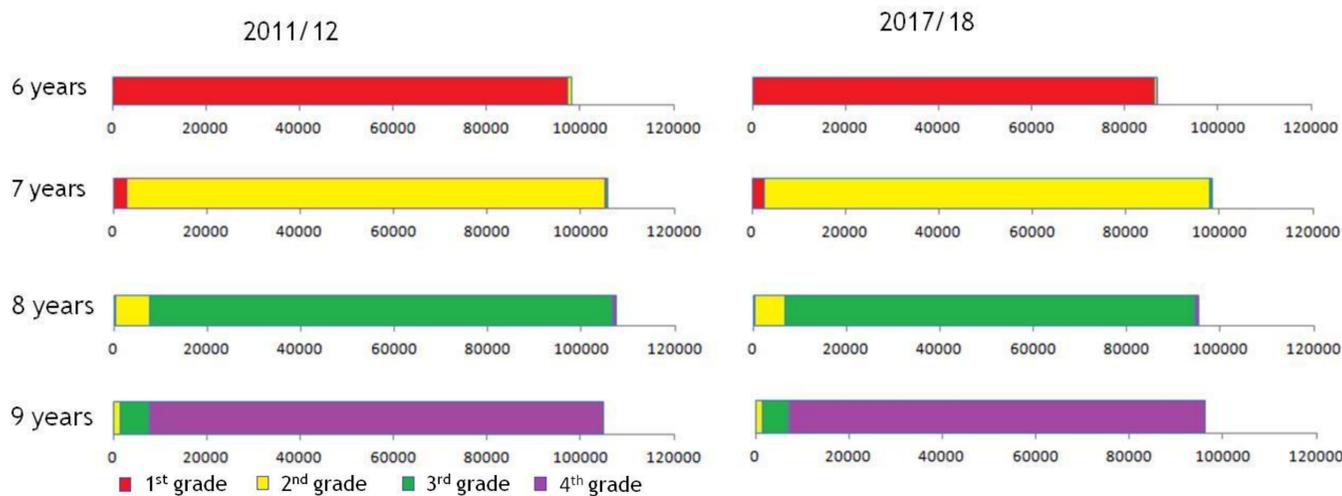$\beta_{X,i}^{(a)}$ - dropout rate at age a, a = 15,16,17 and 18, the i-th school grade of the training offer X.
$c_t^{(a)}$ - proportion of students with a years old in year t, are covered by the extension of compulsory schooling.

Aplications

The sustainable planning of financial, technical and human resources needed for futures school years.
The adaptation of the model to regional realities, useful for analysis and design of the Portuguese school network.

2011/12   2017/18

6 years / 7 years / 8 years / 9 years

■ 1st grade ■ 2nd grade ■ 3rd grade ■ 4th grade

STUDENTS ENROLLED IN THE 1st GRADE OF PRIMARY EDUCATION
– Predictions until 2017/18 –
102,016; 98,493; 95,769; 97,649; 93,237; 94,780; 90,545

STUDENTS ENROLLED IN THE 2nd GRADE OF PRIMARY EDUCATION
– Predictions until 2017/18 –
112,426; 112,293; 108,572; 105,566; 107,400; 102,823; 104,271

STUDENTS ENROLLED IN THE 3rd GRADE OF PRIMARY EDUCATION
– Predictions until 2017/18 –
109,210; 106,419; 106,292; 102,755; 99,950; 101,650; 97,703

STUDENTS ENROLLED IN THE 4th GRADE OF PRIMARY EDUCATION
– Predictions until 2017/18 –
111,477; 109,834; 107,124; 106,916; 101,464; 100,671; 102,195

STUDENTS ENROLLED IN THE 1st CYCLE OF PRIMARY EDUCATION
– Predictions until 2017/18 –
435,129; 427,039; 417,757; 412,886; 404,051; 399,924; 394,346

STUDENTS ENROLLED IN THE 1st CYCLE OF PRIMARY EDUCATION
– School year 2011/12 = 100; Predictions until 2017/18 –
100.0; 98.1; 96.0; 94.9; 92.9; 91.9; 90.6

Metodology

The development of the model is divided in two distinct parts:

✓ estimation of the probabilities of a given student being in grade i with age a. In that sense, it was created a function in R, which was named *probabilidades* and which calculates those estimates.
✓ creation of a matrix containing the predictions obtained through the presented models. In this matrix, the entry for the line i of the column j represents the predicted number of students aged i and attending grade j.

The whole procedure is developed with the function *previsões* which operates by introducing the following arguments:
❖ *X* – The matrix with the most recent data regarding the number of enrollments by age and grade.
❖ *Y* – Time series with the number of births until the most recent year.
❖ *Idadeinicial* – The minimum age for which the prediction is intended.
❖ *Idadefinal* – The maximum age for which the prediction is intended.
❖ *Anoinicial* – The lowest grade for which the prediction is intended.
❖ *AnoFinal* – The highest grade for which the prediction is intended.
❖ *Anolectivo* – School year for which the prediction is intended.
❖ *Anocivil* – The (current) calendar year at the time the prediction is made.

Future developments

Application of the model to other course types.
Variation of factors currently seen as constraints, such as graduation rates, enrolment age structure.
To develop the model in order to be able to predict the number of enrolled students by municipality in Portugal.

MAIN FINDINGS

The forecasts exercise suggests that the increase in the number of students expected from the extension of compulsory education is fully absorbed by the expected demographic reduction.
Consequently, in what concerns compulsory education, the main change will be noticed in the relation between the number of students enrolled in primary an lower secondary education vs. upper secondary education.