Modeling Literacy as a Function of Education and Age

Antero Malin
Institute for Educational Research, University of Jyvaskyla
Yliopistonkatu 9
FIN-40100 Jyvaskyla, Finland
antero.malin@ktl.jyu.fi

1. Literacy, Education and Age

The International Adult Literacy Survey was conducted in 22 countries in 1994-1998. In this survey, literacy was not examined as a basic, mechanical reading skill, but as an adult skill of *"using printed and written information to function in society, to achieve one's goals, and to develop one's knowledge and potential"*. The range of adults' literacy was assessed in three domains: in *prose literacy*, *document literacy*, and *quantitative literacy*. Literacy was measured on three scales ranging from 0 to 500 points.

Among the various background factors explaining literacy skills of the adult population, the most important ones were initial formal education and the respondent's age. In all countries adults with more education had higher average literacy scores. And the higher the age of the respondent, other variables held constant, the lower was the level of literacy on average.

The aim of the paper is, by means of the polynomial regression model, to analyse more accurately the relationship between literacy and its two important determinants. As an example, the model is fitted to the Finnish adult literacy data. In the analysis, education is measured as the years of formal education completed, ranging from 6 to 20 years, and the respondent's age ranges from 16 to 65 years.

2. Polynomial Regression Model

The relationship between literacy and education is often thought to be linear. This means that one additional year in education increases the literacy proficiency evenly, independently of the amount of earlier education. If the relationship between literacy and age is linear, we believe that the mean differences in literacy proficiency between age cohorts are equal as well. But more thorough analysis has shown that straight-line relationships provide a poor fit to the data.

If a linear relationship provides a poor fit, one solution to improve the fit of the statistical model to the data is to use functions of higher powers of the explanatory variables, that is, polynomials of order higher than one. In addition to the linear effects, we then use higher order polynomials in the model to describe the curvilinear effects of education and age. In the Finnish data, modeling literacy as a function of education and age needs both the second and third order effects of age and the second order effect of education. In addition, the interaction of the linear effects is needed for modeling prose literacy. The statistical model used to explain prose literacy as a curvilinear function of the respondent's age and education is the polynomial regression model

$$(1) \qquad y=\beta_0+\beta_1 x_1+\beta_2 x_1^2+\beta_3 x_1^3+\beta_4 x_2+\beta_5 x_2^2+\beta_6 x_1 x_2+\varepsilon$$

where y is the literacy score, x_1 = age - 16 and x_2 = education - 6. The model explains 38 % of the variation in prose literacy. The model used to explain quantitative literacy is model (1) without the interaction $x_1 x_2$, and it explains 26 % of the variation in quantitative literacy.

Figure 1. Prose literacy as a function of education in selected age cohorts in Finland

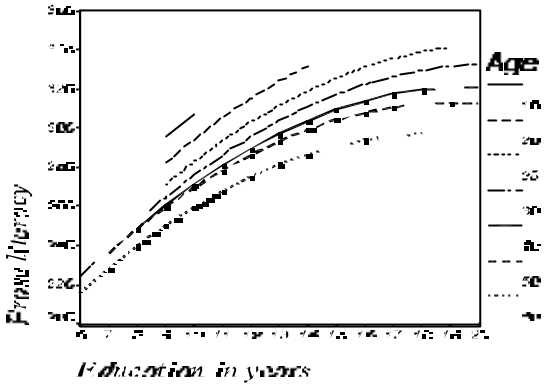
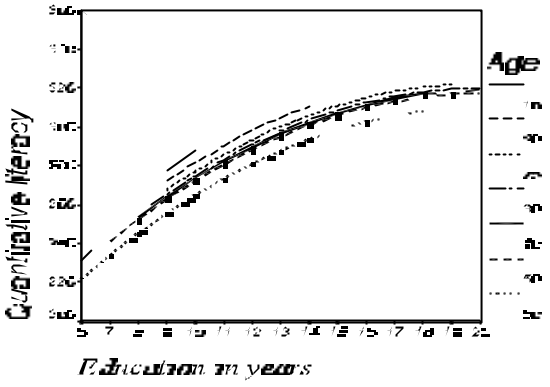


Figure 2. Quantitative literacy as a function of education in selected age cohorts in Finland



3. Results from the Finnish Data

Results of fitting the polynomial regression models to the Finnish data can be seen from Figures 1 and 2, which describe the relationship of education to prose and quantitative literacy in selected age cohorts.

In Figure 1 we can see that more education means better prose literacy in all age cohorts. The relationship is not linear. Increasing education has strongest effect among those with short formal education, while increasing education has almost no effect among those with long education. The differences between age cohorts are larger among those with long education.

In Figure 2 we can see that more education also means better quantitative literacy in all age cohorts, although education has a smaller effect on quantitative literacy than on prose literacy. Increasing education has again strongest effect among those with short education, but the mean differences between age cohorts remain the same, independently of the amount of education. In addition, the differences between age cohorts are smaller than in prose literacy.

The best literacy proficiency is associated with long education and young age. The longer the education, the better the literacy proficiency, compared with people of the same age with shorter education. Since the youngest respondents have not completed their education yet, we have to control the effect of education in comparing the literacy proficiency of different age cohorts. Then we can see that the younger the respondent is, the better is his or her literacy proficiency on average.

The use of the polynomial regression model in modeling literacy as a function of age and education is further illustrated with examples from other countries in the International Adult Literacy Survey.

REFERENCES

Aczel, A. D. 1999. Complete Business Statistics. New York: Irwin/McGraw-Hill.

Draper, N. R. & Smith, H. 1966. Applied Regression Analysis. New York: John Wiley & Sons, Inc.

Literacy in the Information Age. Final Report of the International Adult Literacy Survey. 2000. Pariisi/Ottawa: OECD & Statistics Canada.

Literacy Skills for the Knowledge Society. Further results from the International Adult Literacy Survey. 1997. Pariisi/Ottawa: OECD, HRDC & Statistics Canada.

RESUME

Le modèle de règression polynomiale est utilis pour dècrire les effets curvilinèaires de l'èducation et de l'âge sur la littératie des adultes. Les exemples proviennent de l'Enquête internationale sur la littératie des adultes.