
Socioeconomic Status of Families [nels]

Assignment

Problem

Find out which characteristics are associated with socio-economic status of the family and describe the association.

Population

The data come from the “*National Education Longitudinal Study*” conducted by the U.S. Department of Education. Each record represents a family having a child in the 8th grade of the elementary school in 1988.

Specifications

- (i) Investigate the association of socio-economic status of the family (SES quartile) with the achieved level of education of the father. Use standard methods for the analysis of the two-way contingency table first, then fit a loglinear model. Compare the results. In the loglinear model fit, choose one of the main effects and one of the interaction terms and interpret the estimated parameters.
- (ii) Investigate the association of socio-economic status (SES quartile) with father’s education and region of residence by a loglinear model. Is the association of socio-economic status with father’s education the same in every region? If not, describe how it varies in different regions. In which region is the dependence strongest?
- (iii) Investigate the association of socio-economic status (SES below/above median) with father’s education, father’s work status, and region of residence by a logistic regression model fit on grouped data.
- (iv) Investigate the association of socio-economic status (SES below/above median) with father’s education, father’s work status, and region of residence by the loglinear model that is equivalent to the logistic model fitted in (iii). Show that the relevant parameter estimates and test results are the same.
- (v) Build a general loglinear model investigating the mutual associations between all the variables in the data set. For socioeconomic status, use the two-level variant (below/above median), not the SES quartiles. Take care not to fit models with too many parameters (some of the models have thousands of parameters and take too long to fit). Interpret the final model – what does it imply about conditional independence of the variables? Choose one of the main effects and one of the two-way interaction terms in the final model and interpret the estimated parameters.

Requirements

Write a report (prepared by \LaTeX , LibreOffice, MS Word, ...) summarizing your solution to the problems specified during the exercise classes.

More precise specification of what exactly should be (and also what should not be) included in the report will be provided during the exercise classes related to this assignment (April 11, 18, 25). Pay attention to those instructions!

The report in the pdf format (file named as Surname_Firstname_3.pdf) and the related R script (file named as Surname_Firstname_3.R) have to be submitted in Moodle by **Sunday April 30, 2023 [23:59 CEST]**.

Dataset

The dataset can be downloaded from

http://www2.karlin.mff.cuni.cz/~komarek/vyuka/2022_23/nmst412/Problem_3/GLM_3_nels.RData

The dataframe is called `nels`. It contains 13 580 rows (families) and 9 variables.

Variable list: See Table 1.

Table 1: Variable coding table

Variable Name	Variable Label	Variable Coding
<code>ses</code>	Socio-economic status (quartile)	1, 2, 3, 4
<code>sesmed</code>	Socio-economic status (median)	below, over
<code>parents</code>	Number of adults in family	1, 2
<code>foreign</code>	Foreign language spoken at home	yes, no
<code>fa.educ</code>	Father's education	factor
<code>mo.educ</code>	Mother's education	factor
<code>region</code>	Region within US	factor
<code>fa.wrk</code>	Father working	yes, no
<code>mo.wrk</code>	Mother working	yes, no