

Week 5 Lab Exercises

Philip Leifeld

GV903 Advanced Methods – University of Essex, Department of Government

1 Typesetting and Integration with R

1. Inspect the `.tex` source file for the exercises and solutions in all previous weeks, and learn \LaTeX . Install a \LaTeX distribution on your computer (e. g., \TeX or \TeX Live), download a \LaTeX editor, and typeset a document with the solutions for this set of exercises in \LaTeX .
2. Familiarise yourself with Markdown, `knitr`, and the way these two functionalities can be used within RStudio to integrate R code within Markdown documents. Typeset your solutions for this Week’s exercises in Markdown and make sure you embed some R code into the document.
3. Familiarise yourself with `.Rnw` documents in RStudio. This is a way to embed R code also into \LaTeX code and compile the integrated results into a PDF file. Inspect and compile the `.Rnw` documents from the previous weeks. Also create a solution document for the tasks in this document by integrating R code into your \LaTeX document using RStudio, `knitr`, and a `.Rnw` document.

I can recommend some additional online resources for learning Markdown (#1, #2), `knitr` with Markdown (#3, #4), #5), \LaTeX (#6, #7, #8), and `knitr` with \LaTeX (#9). I will also post an additional PDF document with details about \LaTeX on Moodle.

2 Linear model of pupils’ verbal expression

Table 1 contains some fictitious observations for 20 different schools. The `Salary` variable measures how many thousand Pound Sterling in teacher salary are spent per pupil and year. The `WhiteCollar` variable is the percentage of fathers with white-collar jobs of the pupils in each school. `SES` measures the average socio-economic status of all pupils in the respective school. `TeachScore` reports the mean of a verbal test score for all teachers per school. `MotherEduc` is the mean educational level of the pupils’ mothers, where a value of 1 corresponds to two years, 2 corresponds to four years etc. `VerbalScore` measures the verbal test score of all pupils in the respective school.

1. In R, write code to create this data frame without relying on any external files or interactive functions (e. g., graphical windows) for data input. Create a scatterplot for the relationship between the average educational level of the mother and the share of white-collar fathers. Show the R code for both, and show the scatterplot.
2. Write down the mathematical equation for a linear regression model in which the verbal test score of pupils is regressed on the remaining variables.

Salary	WhiteCollar	SES	TeachScore	MotherEduc	VerbalScore
3.83	28.87	7.20	26.60	6.19	37.01
2.89	20.10	-11.71	24.40	5.17	26.51
2.86	69.05	12.32	25.70	7.04	36.51
2.92	65.40	14.28	25.70	7.10	40.70
3.06	29.59	6.31	25.40	6.15	37.10
2.07	44.82	6.16	21.60	6.41	33.90
2.52	77.37	12.70	24.90	6.86	41.80
2.45	24.67	-0.17	25.01	5.78	33.40
3.13	65.01	9.85	26.60	6.51	41.01
2.44	9.99	-0.05	28.01	5.57	37.20
2.09	12.20	-12.86	23.51	5.62	23.30
2.52	22.55	0.92	23.60	5.34	35.20
2.22	14.30	4.77	24.51	5.80	34.90
2.67	31.79	-0.96	25.80	6.19	33.10
2.71	11.60	-16.04	25.20	5.62	22.70
3.14	68.47	10.62	25.01	6.94	39.70
3.54	42.64	2.66	25.01	6.33	31.80
2.52	16.70	-10.99	24.80	6.01	31.70
2.68	86.27	15.03	25.51	7.51	43.10
2.37	76.73	12.77	24.51	6.96	41.01

Table 1: Verbal expression data

3. In R, estimate this equation using ordinary least squares. You can use existing functions to do so; no need to re-implement OLS. Interpret the effects of **Salary**, **SES**, and **TeachScore** on **VerbalScore** in terms of effect size, uncertainty, and significance. Show the regression table using the **texreg** package.
4. Interpret the goodness-of-fit measures (R^2 and F test) in up to 150 words.
5. Show with the empirical data how R^2 and the F test are calculated, both using equations in which you insert actual values and R code. You can take the residuals and/or predicted values from the model you estimated before.