

# **EFECTAS: R Tool for International** Large-Scale Assessment Data Analysis Laura Ringienė, Audronė Jakaitienė



The quality of education is an issue around the world. Researchers analyse not only national education data but also cross-country comparisons. High volume and special structure International Large-Scale Assessment data (ILSA) such as PISA (Programme for International Student Assessment), TIMSS (Trends in International Mathematics and Science Study), PIRLS (Progress in International Reading Literacy Study), and others, are used to compare students' achievement in mathematics, science, and reading within and between countries. Such data can be analysed using commercial software such as SPSS, SAS, Mplus, etc., or the open source software R. We reviewed five open-source R software packages for statistical analysis of ILSA data: BIFIEsurvey, EdSurvey, intsvy, RALSA, and svyPVpack (Ringiene, Žilinskas and Jakaitiene, 2021, Jakaitiene and others, 2022). After analysing the advantages and disadvantages of each package, we have developed a new tool, EFECTAS, based on the BIFIEsurvey and EdSurvey packages.



## https://www.efectas.projektas.vu.lt/index.php/en/software

function result <- function name(parameter1, parameter2, ...)</pre> All EFECTAS functions have the same structure Summary(function result) The majority of results are presented in Descriptive statistic informative on-screen tables and excel documents. Frequency table Percentile Categorical and continuous variables cannot be used at the same Download and prepare time. Descriptive statistics are also provided for grouped variables. Only fo one continuous variable The data is downloaded from the Categorical variables Continuous variables Call: percentile(variable = variables, percentiles = percent, data = Pisa.data, weightVar = "w\_fstuwt") official website using an internet full data n: 6525 On screen On screen n used: 6334 connection. FREQUENCY TABLE df confInt.ci\_lower confInt.ci\_upper nsmall estimate se variable variable value N Weighted N All Percentage Full Percentage 5 -1.43910118 0.02301012 21.08789 -1.3142000/ariable st011q12ta 50 5625 1115 95 3377 3508 364 1708 2806 1214 598 195 1328 3667 1890 177.0 Weighted N Na's Median 3rd Ou. 25 -0.78086288 0.03868594 21.1592 0.5099690 1600 177.05 1985.66 4090.78 299.78 11958.90 12494.37 1144.57 6372.53 9770.18 4253.29 2232.38 680.33 4938.54 12907.16 6607.57 st011q12ta st011q12ta st011q12ta st004d01t 6% downloaded 81.32 16.73 1.23 48.91 51.09 4.68 26.06 39.95 17.39 9.13 2.78 20.20 52.78 27.02 82.94 17.06 
 481.19
 91.39
 155.21
 418.26
 482.86

 475.87
 94.30
 154.22
 409.65
 478.93

 0.03
 0.87
 -6.22
 -0.66
 0.13

 53.08
 23.24
 11.74
 29.16
 58.07
544.63 543.24 0.74 73.91 810.70 775.52 3.79 88.96 24453.27 0 24453.27 0 23819.75 192 21692.52 827 6885 6885 6693 6058 math read escs hisei 50 0.02421431 0.04347068 31.0977 -0.3308272 0.3452626 3105 NO RESPONSE FEMALE MALE 0.8688038 1556 48.91 51.09 75 0.66990000 0.01804455 24.00991 0.4492029 URL: http://webfs.oecd.org/pisa/PUF\_SPSS\_COMBINED\_CMB\_STU\_QQQ.zip st004d01t 95 1.14122770 0.02074189 32.36995 1.0071749 1.4130047 st034q02ta (Missing) 28.16 43.18 18.80 9.87 st034q02ta STRONGLY AGREE st034q02ta AGREE st034q02ta DISAGREE st034002ta STRONGLY DISAGREE st034q02ta st034q02ta sc012q01ta sc012q01ta sc012q01ta NO RESPONSE ALWAYS NEVER 20.20 52.78 27.02 Benchmarks Preparing data for analysis: Excel Exce Merge data files Weighted N Na's Mean SD entage | Ful Only fo one variable consisting of 50 177.05 0 481.19 91.39 155.21 418.26 482.86 544.63 810.7 st011q12ta (Mis 0.72 math 6885 24453.27 5625 81.32 st011q12ta YES 19885.66 82.94 24453.27 0 475.87 94.3 154.22 409.65 478.93 543.24 775.52 read 6885 Select variables 1115 4090.78 16.73 17.06 st011q12ta NO plausible values 23819.75 192 0.03 0.87 -6.22 -0.66 0.13 0.74 3.79 escs 6693 299.78 1.23 st011g12ta NO RESPONSE 95 21692.52 827 53.08 23.24 11.74 29.16 58.07 73.91 88.96 hisei 6058 48.91 48.91 t004d01t FEMAL 11958.9 Recode data 51.09 51.09 t004d01t 3508 12494.37 364 4.68 t034a02ta (Missing) 1144.57 1708 26.06 28.16 t034q02ta STRONGLY AGREE 6372.53 39.95 43.18 t034g02ta AGREE 2806 9770.18 17.39 18.8 t034q02ta DISAGRE 1214 4253.29 9.87 9.13 st034g02ta STRONGLY DISAGREE 598 2232.38 0.73 st034q02ta NO RESPONSE 680.33 2.78 195 0.57 6.74 1328 4938.54 20.2 20.2 c012q01ta ALWAYS 1.35 0.29 12907.16 52.78 52.78 3667 c012q01ta NEVER Regression 27.02 27.02 12q01ta SOMETIME 1890 6607.57 One dependent variable and for several independent variables Two level multilevel modeling Formula: scie ~ st011q12ta + st004d01t

One model

On screen

Variance method: jackknife JK replicates: 80 Plausible values: 10 jrrIMax: 1 full data n: 6525 n used: 6254

Weight variable: 'w fstuwt'

#### Coefficients:

t dof Pr(>|t|) coef se (Intercept) 482.6121 2.9923 161.2830 65.385 < 2.2e-16 \*\*\* st011q12taN0 -22.4961 5.3158 -4.2319 69.994 6.911e-05 \*\*\* st004d01tMALE -5.0634 3.0436 -1.6636 64.623 0.101

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Multiple R-squared: 0.0074

### Reference

Ringienė, Laura, Žilinskas, Julius, & Jakaitienė, Audronė (2021). ILSA data analysis with R packages. *Modelling, Computation* and Optimization in Information Systems and Management Sciences : Proceedings of the 4th International Conference on Modelling, Computation and Optimization in Information *Systems and Management Sciences - MCO 2021, 271–282.* 

Jakaitienė, Audronė, Želvys, Rimantas, Raižienė, Saulė, Dukynaitė, Rita, Stumbrienė, Dovilė, Vaitekaitis, Jogaila, Ringienė, Laura, Žilinskas, Julius, & Žilinskas, Antanas. (2022). (Ne)išmatuojamas švietimo efektyvumas ir našumas. Vilniaus universiteto leidykla.



### Correlation

#### Pearson correlation.

Between continuous variables and plausible values. Can be calculated by group.

2	Correlat	tion sta	TISTIC									
	var1	var 2	Ncases	Nweight	cor	cor_SE	t	df	p	cor_fmi	cor_VarMI	cor_Var
2	escs	pv1scie	6334	29037.19	0.3423767 0.	018891263	18.12	Inf	2.215950e-73	0	0	3.568798e
3	escs	pv2scie	6334	29037.19	0.3470641 0.	018183736	19.09	Inf	3.057694e-81	0	0	3.306482e
5	pv1scie	pv2scie	6334	29037.19	0.9136037 0.	002796806 3	26.66	Inf	0.000000e+00	0	0	7.822124e
s	Correlat	tion mat	rix`									
S	Correlat	tion mat	rix`\$on	e1								
		escs	pv1s	cie pv2	scie							
e	scs 1.	. 0000000	0.3423	767 0.347	0641							
p	viscie 0.	. 3423767	1.0000	000 0.913	6037							
p	v2scie 0.	.3470641	0.9136	037 1.000	0000							
S	Covaria	nce stat	istic'									
1	var1	var2	Ncases	Nweight	COV	cov_S	E COV_	df	cov_VarRep			
1	escs	escs	6334	29037.19	0.7532142	0.014980	7 I	nf	2.244215e-04			
2	escs	pv1scie	6334	29037.19	26.8599562	1.764040	6 I	nf	3.111839e+00			
3	escs	pv2scie	6334	29037.19	27,6094959	1.750133	6 I	nf	3.062968e+00			
4	pv1scie	pv1scie	6334	29037.19	8171, 1609019	240.110934	1 I	nf	5.765326e+04			
5	pv1scie	pv2scie	6334	29037.19	7569,8851817	242.894026	7 I	nf	5.899751e+04			
6	pv2scie	pv2scie	6334	29037.19	8401,931248	3 258,173248	6 I	nf	6.665343e+04			

\$`Covariance matrix` \$`Covariance matrix`\$one1 escs pv1scie pv2scie 0.7532142 26.85996 27.6095 pv1scie 26.8599562 8171.16090 7569.8852 /2scie 27.6094959 7569.88518 8401.9312

MODEL INFO:				
Plausible Valu	es: 5			
Number of Grou	ps:			
Level Group	n size	mean wgt	sum wgt	
2 idscho 1 Obs	ol 207 3741	3.95 5.94	817.22 22231.49	
Formula: mmat	~ 1 + (1   id	school)		
Intraclass Cor	relation= 0.3	92		
MODEL FIT INFO	RMATION:			
Loglikelihood AIC BIC	Mean -124977.4 249960.8 249979.5	Std Dev 149.31 298.62 298.62		
MODEL RESULTS:  Fixed Effects:				
	Estimate S	td. Error	t value	P-value
(Intercept)	523.42	6.54	80.09	0

#### Std. Error t value imate Observation 3741 523.42 4076.50 113.12 ariance/Residual variance/ 2631.38 608.01 4.33 249960.7 249979.4

Random Effects: Variance Std Error Std Dev

	di oup	iveance.	var rance	500. 20100	bea.bev.	c varac	i varac	
2	idschool	(Intercept)	2631.38	608.01	51.30	4.33	0	
1	Residual		4076.50	113.12	63.85	36.04	0	

Model data saved in a document multilevel.xlsx

#### Model comparison

Excel

	Model 1				Model 2				Model 3			
	Estimate	Std. Error	t value	P-value	Estimate	Std. Error	t value	P-value	Estimate	Std. Error	t value	P-value
Observation	3741				3032				3032			
Intercept	523.42	6.54	80.09	0.00	577.19	7.65	75.46	0.00	575.40	6.90	83.36	0.76
Within level												
asdhedup					-26.40	2.35	-11.23	0.00	-26.29	2.35	-11.21	0.82
Variance/Residual variance	4076.50	113.12	36.04	0.00	3814.55	131.11	29.09	0.00	3812.64	131.75	28.94	0.79
Between level												
asdhaps2l									4551.24	1849.24	2.46	0.88
Variance/Residual variance	2631.38	608.01	4.33	0.00	1757.76	511.62	3.44	0.01	23225.71	9363.94	2.48	0.88
AIC	249960.78				202316.42				202215.43			
BIC	249979.46				202340.49				202251.53			
ICC	0.39				0.32				0.88			
The best model according to AIC is Model 3												
The best model according to BIC is Model 3												

Excel

+ value \_\_value