



ACRO

Automated checking of research output

*The 2021 joint UNECE/Eurostat Expert Meeting
on Statistical Data Confidentiality
1-3 December 2021.*



Introduction

- *Pilot project commissioned by Eurostat in 2019*
- *Delivered in 2021 by UWE* – Bristol*
- *Proof-of-concept software based on STATA®*
- *Designed as a tool capable to reduce the workload of output checking, traditionally made manually by output checking officers*
- *Published as Statistical Working Paper "Automatic Checking of Research Outputs (ACRO): a tool for dynamic disclosure checks"*
- *Source code published on <https://github.com/eurostat/ACRO>*

* F.Ritchie, J.Smith and E.Green

Functional features

- 1. Implementation of automatic checks on:*
 - **Tabulations**
 - **Common estimators**
 - **Medians**
 - **Maxima/Minima (simple ban)**
- 2. Automatic primary suppression*
- 3. Requests of exceptions*
- 4. Report preparation (Excel)*

Statistical Disclosure rules applied

Rule	Description	Applies to	Commands
<i>Threshold</i>	Minimum number of observations underlying a statistic	All linear statistics: frequencies, mean, median, sums etc	table tabulate
<i>N-K dominance</i>	The N largest observations should not count for more than K% of the total	As for threshold, but doesn't apply to frequencies	table tabulate
<i>P-ratio</i>	It should not be possible to estimate another observation within p% of its value	As for N-K rule	table tabulate
<i>Table rule</i>	SDC rules are applied to each table cell independently; any cell can pass or fail	Applies to all tables	table tabulate
<i>Maximum & minimum</i>	Not allowed	Any magnitude	table tabulate
<i>Degrees of freedom</i>	Analytical outputs must have at least K degrees of freedom	Analytical results, including estimation and testing	regress, xtreg logit, probit test, ttest

Features not implemented

- 1. Automatic checks on graphs (all set to review)*
- 2. Percentiles (except median)*
- 3. Secondary (within table) suppression*

Practical use

A researcher:

- *starts the STATA[®] platform, imports the data*
- *runs a '**setup**' script*
- *uses the keyword '**safe**' to prefix the commands*
- *uses '**finalise**' to prepare the spreadsheet for review by the output checker.*

Sample ACRO output

test_results.xlsx - Ex

File Home Insert Page Layout Formulas Data Review View Developer ACROBAT Power Pivot Tell me what you want to do...

A2 :

	A	B	C	D	E	F	G
	Sheet	Automatic check	Final decision	Description safe/unsafe	Reason for automatic decision	Exception request	Additional notes
1	activity	ok	ok	unsafe statistic: table	pass	n/a	
2	graph test	review		graph: twoway	review required		
3	max act	ok	ok	unsafe statistic: table	fail; suppression app	n/a	
4	output 1	ok	ok	unsafe statistic: tabulate	pass	n/a	
5	output 2 A	fail	fail	unsafe statistic: tabulate	fail	n/a	
6	output 2 B	ok	ok	unsafe statistic: tabulate	pass	n/a	
7	output 2 C	ok	ok	unsafe statistic: tabulate	pass	n/a	
8	output 2 D	fail	fail	unsafe statistic: tabulate	fail	n/a	
9	output 3	fail	fail	unsafe statistic: table	fail	n/a	
0	output 4	review		unsafe statistic: table	fail; exception requ	trust me, I'm a professor	
1	output 5	ok	ok	safe statistic: regress	pass	n/a	
2	small act A	review		unsafe statistic: table	fail; exception requ	It's not feasible to identify the charities from this information	
3	small act B	review		unsafe statistic: table	fail; exception requ	It's not feasible to identify the charities from this information	
4	small act C	review		unsafe statistic: table	fail; exception requ	It's not feasible to identify the charities from this information	



European
Commission



Advantages

- *Overall good efficiency*
- *Minimal setup required*
- *Low training overhead*
- *Readability of output results (MS Excel[®] files)*

Limitations

- *Only one language available (Stata scripts)*
- *Narrow range of implemented functions*

Initial review

Initial reviewers find that ACRO:*

- *has good potential to reduce output checking workload*
- *is not perfectly intuitive as expected, though it can be learnt with a minimal effort*
- *implementing a wider range of functions would substantially improve the utility of the tool*
- *porting the tool to a wider set of languages (R/SAS) is necessary*

(*) SDC experts and managers of RDCs

Next steps

- *We welcome feedback in order to guide next developments:*
 - Feedback by email: estat-confidentiality@ec.europa.eu
 - Feedback by Git: fork the project, work your branch, then pull request
- *Further developments to address limitations and improve general functionalities of the tool and its buy-in by RDC managers and users.*

References

- F.Ritchie, E.Green, J.Smith: "*Automatic Checking of Research Outputs (ACRO): a tool for dynamic disclosure checks*" (*). Eurostat, Statistical Working Papers series. Publication Office of the European Union, Luxembourg (2021).

(*) <https://ec.europa.eu/eurostat/product?code=KS-TC-21-005>



ACRO

Automated checking of research output

Thank you for your attention!

*The 2021 joint UNECE/Eurostat Expert
Meeting on Statistical Data Confidentiality
1-3 December 2021.*

