Statistical Programming and Open Science Methods

Explore your researcher degrees of freedom

Joachim Gassen
Humboldt-Universität zu Berlin

September 02, 2022

SFB/Transregio 266
ACCOUNTING FOR TRANSPARENCY

Time table Tuesday, February 18

When?	What?
09:00	Explore your researcher degrees of freedom
10:30	Coffee
11:00	Providing data access via RESTful APIs
12:30	Lunch
13:30	Group Work Presentations
15:00	Coffee and Wrap Up
15:30	End of Event

Motivation

What troubles me about using opinions is their whimsical nature. Some mornings when I arise, I have the opinion that Raisin Bran is better than eggs. By the time I get to the kitchen, I may well decide on eggs, or oatmeal. I usually do recall that the sixteenth president distinguished himself. Sometimes I think he was Jackson; often I think he was Lincoln.

A data analysis is similar. Sometimes I take the error terms to be correlated, sometimes uncorrelated; sometimes normal and sometimes nonnormal; sometimes I include observations from the decade of the fifties, sometimes I exclude them; sometimes the equation is linear and sometimes nonlinear; sometimes I control for variable z, sometimes I don't. Does it depend on what I had for breakfast?

Leamer (AER 1983: 37f.)

Slightly more systematic ...

TABLE 1 | Checklist for different types of researcher degrees of freedom in the planning, executing, analyzing, and reporting of psychological studies.

Code	Related	Type of degrees of freedom
Hypothesizing		
T1	R6	Conducting explorative research without any hypothesis
T2		Studying a vague hypothesis that fails to specify the direction of the effect
Design		
D1	A8	Creating multiple manipulated independent variables and conditions
D2	A10	Measuring additional variables that can later be selected as covariates, independent variables, mediators, or moderators
D3	A5	Measuring the same dependent variable in several alternative ways
D4	A7	Measuring additional constructs that could potentially act as primary outcomes
D5	A12	Measuring additional variables that enable later exclusion of participants from the analyses (e.g., awareness or manipulation checks)
D6		Failing to conduct a well-founded power analysis
D7	C4	Failing to specify the sampling plan and allowing for running (multiple) small studies
Collection		
C1		Failing to randomly assign participants to conditions
C2		Insufficient blinding of participants and/or experimenters
C3		Correcting, coding, or discarding data during data collection in a non-blinded manner
C4	D7	Determining the data collection stopping rule on the basis of desired results or intermediate significance testing
Analyses		
A1		Choosing between different options of dealing with incomplete or missing data on *ad hoc* grounds
A2		Specifying pre-processing of data (e.g., cleaning, normalization, smoothing, motion correction) in an *ad hoc* manner
A3		Deciding how to deal with violations of statistical assumptions in an *ad hoc* manner
A4		Deciding on how to deal with outliers in an *ad hoc* manner
A5	D3	Selecting the dependent variable out of several alternative measures of the same construct
A6		Trying out different ways to score the chosen primary dependent variable
A7	D4	Selecting another construct as the primary outcome
A8	D1	Selecting independent variables out of a set of manipulated independent variables
A9	D1	Operationalizing manipulated independent variables in different ways (e.g., by discarding or combining levels of factors)
A10	D2	Choosing to include different measured variables as covariates, independent variables, mediators, or moderators
A11		Operationalizing non-manipulated independent variables in different ways
A12	D5	Using alternative inclusion and exclusion criteria got selecting participants in analyses
A13		Choosing between different statistical models
A14		Choosing the estimation method, software package, and computation of SEs
A15		Choosing inference criteria (e.g., Bayes factors, alpha level, sidedness of the test, corrections for multiple testing)
Reporting		
R1		Failing to assure reproducibility (verifying the data collection and data analysis)
R2		Failing to enable replication (re-running of the study)
R3		Failing to mention, misrepresenting, or misidentifying the study preregistration
R4		Failing to report so-called “failed studies” that were originally deemed relevant to the research question
R5		Misreporting results and *p*-values
R6	T1	Presenting exploratory analyses as confirmatory (HARKing)

Wicherts et al (Frontiers in Psychology 2016: 3)

Objective

Development of a statistical computing framework for a "multiverse analysis" (Steeger et al., Persp on Psych Sci 2016) that

- supports research design development separate from data,
- supports pre-registering of observational data research designs,
- facilitates a priori power analysis,
- promotes unit testing,
- generates well documented and easily portable code,
- makes researcher degrees of freedom (Simmons et al., Psyc Science 2011) explicit in code, and
- allows for rigorous robustness checking by exhausting all these degrees of freedom algorithmically.

Note: Code based walk-through on what follows: https://joachim-gassen.github.io/rdfanalysis/articles/analyzing_rdf.html

Workflow

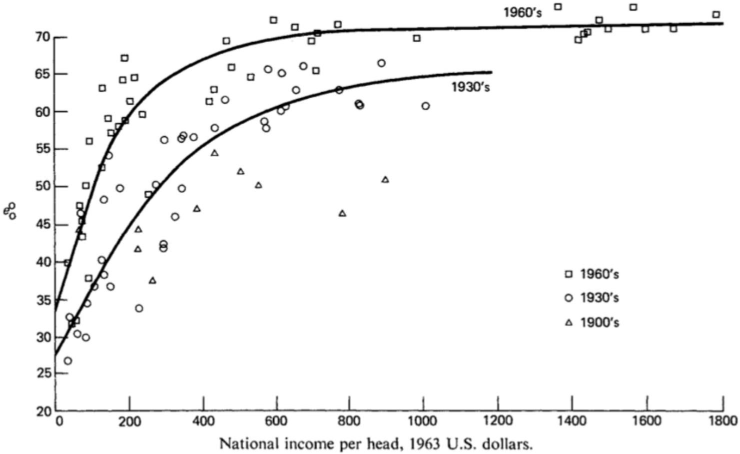
1. Sketch the `design`, the `result` and the required `steps` to generate it
2. Develop test code for `input` data
3. Develop function to simulate `input` data. Verify it with test code
4. For each `step`
   - Draft `step_description` and develop test code
   - Identify and implement potential `choices`. Document them by providing the `choice_description`
   - Verify that `step` passes tests
5. Verify that `design` produces unbiased `results` consistent with simulated effect sizes
6. Run power simulation with your a priori estimation of effect size
7. Register your priors about choices by assigning `weights`
8. Run `design` with real data and interpret `result`

Implementation of a `design`

- A `design` consists of `steps`.
- Each `step` reads the `output` of the prior `step` as `input` and can contain any number of
  - `discrete choices` and
  - `continuous choices`
- The output of the final `step` provides the `result`
- By referring to the `protocol` of prior `choices`, this set-up can be used to design arbitrarily complex forking paths (Gelman and Loken, Am Scientist 2011)

A case study: Preston (1975)

Scatter-diagram of relations between life expectancy at birth (e_0^0) and national income per head for nations in the 1900s, 1930s, and 1960s.



Preston (1975): The Changing Relation between Mortality and level of Economic Development, Population Studies (29): 235.

Descriptive Statistics

	N	Mean	Std. dev.	Min.	25 %	Median	75 %	Max.
Life expectancy	4,139	68.0	10.0	26.2	61.2	70.6	75.4	84.3
GDP per capita	4,139	11.9	17.3	0.2	1.3	3.9	14.0	112.0
Years of schooling	4,139	7.8	3.0	0.3	5.6	8.4	10.3	13.2
Unemplyoment rate	4,139	8.2	6.4	0.2	3.6	6.5	10.8	37.9

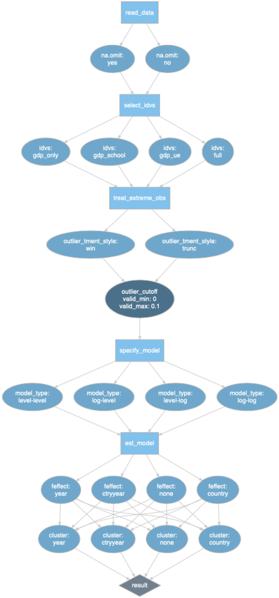
Note: The data is obtained from the World Bank and the Wittgenstein Center. The sample covers 179 countries and the period 1991 to 2015. *GDP per captia* are in constant 2010 thousand U.S. dollars.

Regression Results

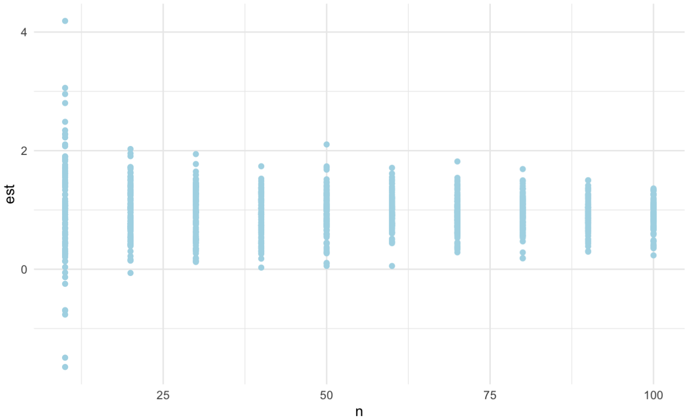
	Dependent variable: *Life expectancy*			
	(1)	(2)	(3)	(4)
ln(*GDP per capita*)	5.211^{***}	3.709^{***}	3.741^{***}	1.499^{***}
	(0.062)	(0.082)	(0.034)	(0.491)
ln(*Years of schooling*)		6.432^{***}	6.006^{***}	6.851^{***}
		(0.241)	(0.081)	(1.689)
ln(*Unemployment rate*)		-0.962^{***}	-0.881^{***}	0.014
		(0.108)	(0.124)	(0.355)
Constant	24.499^{***}	26.227^{***}		
	(0.524)	(0.496)		
Fixed effects	None	None	Year	Country, Year
Std. errors clustered	No	No	Year	Country, Year
Observations	4,139	4,139	4,139	4,139
R^2	0.632	0.687	0.681	0.109
Adjusted R^2	0.632	0.687	0.678	0.063

Note: The dependent variable is the average life expectancy at birth in years. OLS coefficients are reported together with standard errors in parentheses. $^{*}/^{**}/^{***}$ indicate two-sided significance levels of 10/5/1 %, respectively.

The steps and their discrete/continuous choices

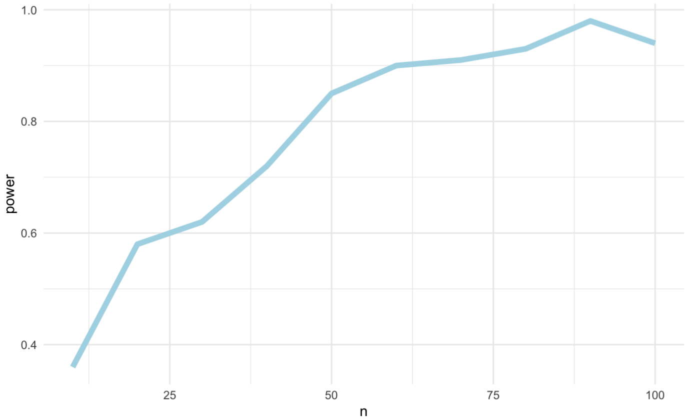


Optional: A power analysis on simulated data I



Power estimation based on an assumed effect size of one year for a 10 increase of GDP per capita. Each country has 15 years of data. n: Number of countries in sample.

Optional: A power analysis on simulated data II



Power estimation based on an assumed effect size of one year for a 10 increase of GDP per capita. Each country has 15 years of data. n: Number of countries in sample.

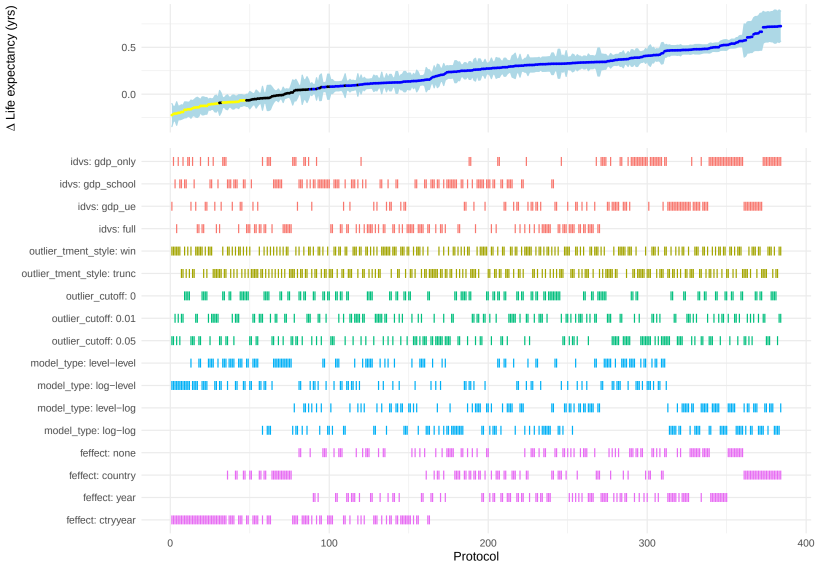
Estimate the effect size on real data

na.omit	idvs	outlier_tment_style	outlier_cutoff	model_type	feffect	cluster	weight	est	lb	ub
yes	full	win	0.00	level-log	ctryyear	ctryyear	0.05	0.136	0.046	0.225
yes	full	trunc	0.00	level-log	ctryyear	ctryyear	0.05	0.136	0.046	0.225
yes	full	win	0.01	level-log	ctryyear	ctryyear	0.10	0.118	0.041	0.196
yes	full	trunc	0.01	level-log	ctryyear	ctryyear	0.10	0.104	0.036	0.173
yes	full	win	0.05	level-log	ctryyear	ctryyear	0.10	0.095	0.024	0.167
yes	full	trunc	0.05	level-log	ctryyear	ctryyear	0.10	0.079	0.006	0.152
yes	full	win	0.00	log-log	ctryyear	ctryyear	0.05	0.164	0.053	0.274
yes	full	trunc	0.00	log-log	ctryyear	ctryyear	0.05	0.164	0.053	0.274
yes	full	win	0.01	log-log	ctryyear	ctryyear	0.10	0.135	0.047	0.223
yes	full	trunc	0.01	log-log	ctryyear	ctryyear	0.10	0.116	0.043	0.190
yes	full	win	0.05	log-log	ctryyear	ctryyear	0.10	0.093	0.013	0.174
yes	full	trunc	0.05	log-log	ctryyear	ctryyear	0.10	0.083	0.007	0.160

And the weighted average estimate is:

est	lb	ub	n
0.112	0.031	0.193	12

Some but not all of the researcher degrees of freedom



And an interactive variant of it

https://jgassen.shinyapps.io/shiny_rdf_spec_curve/

For further info and code

https://joachim-gassen.github.io/rdfanalysis