

How to write a Pre-Analysis Plan

A guide for Political Science

Dalson Figueiredo (UFPE, dalson.figueiredofo@ufpe.br)

Lucas Silva (UFPE, lucas.eosilva@ufpe.br)

April 2, 2016

Transparency?

Definition

property of transmitting light without appreciable scattering so that bodies lying beyond are seen clearly (physics)

Definition

to offer detailed information about all characteristics of the research design and data analysis (science)

WHAT IS A PAP AND WHAT IS IT GOOD FOR?

Definition

a pre-analysis plan is a step by step plan setting out how a researcher will analyze data which is written in advance of them seeing this data (and ideally before collecting it in cases where the researcher is collecting the data (McKenzie, 2012)

Definition

a pre-analysis plan is a detailed outline of the analyses that will be conducted in a study (CHRISTENSEN and SODERBERG, 2015: 26)

WHAT IS A PAP AND WHAT IS IT GOOD FOR?

- The main purpose of PAP is to reduce the incidence of false positive results (SIMMONS, NELSON and SIMONSOHN, 2011)

WHAT IS A PAP AND WHAT IS IT GOOD FOR?

- The main purpose of PAP is to reduce the incidence of false positive results (SIMMONS, NELSON and SIMONSOHN, 2011)
- The goal of PAP is to bind our hands against data mining (Casey, Glennerster, & Miguel, 2012)

WHAT IS A PAP AND WHAT IS IT GOOD FOR?

- The main purpose of PAP is to reduce the incidence of false positive results (SIMMONS, NELSON and SIMONSOHN, 2011)
- The goal of PAP is to bind our hands against data mining (Casey, Glennerster, & Miguel, 2012)
- PAP can be used to reduce researcher degrees of freedom (CHRISTENSEN and SODERBERG, 2015: 26)

10 features of PAP

1 Sample size and sampling strategy

10 features of PAP

- 1 Sample size and sampling strategy
- 2 Variables measurement level

10 features of PAP

- 1 Sample size and sampling strategy
- 2 Variables measurement level
- 3 Hypotheses

10 features of PAP

- 1 Sample size and sampling strategy
- 2 Variables measurement level
- 3 Hypotheses
- 4 Full model (functional form, interactions and specification

10 features of PAP

- 1 Sample size and sampling strategy
- 2 Variables measurement level
- 3 Hypotheses
- 4 Full model (functional form, interactions and specification
- 5 Methods and techniques

10 features of PAP

- 1 Sample size and sampling strategy
- 2 Variables measurement level
- 3 Hypotheses
- 4 Full model (functional form, interactions and specification
- 5 Methods and techniques
- 6 Inclusion, exclusion, transformations and corrections

10 features of PAP

- 1 Sample size and sampling strategy
- 2 Variables measurement level
- 3 Hypotheses
- 4 Full model (functional form, interactions and specification
- 5 Methods and techniques
- 6 Inclusion, exclusion, transformations and corrections
- 7 Source of data

10 features of PAP

- 1 Sample size and sampling strategy
- 2 Variables measurement level
- 3 Hypotheses
- 4 Full model (functional form, interactions and specification
- 5 Methods and techniques
- 6 Inclusion, exclusion, transformations and corrections
- 7 Source of data
- 8 Computational software's

10 features of PAP

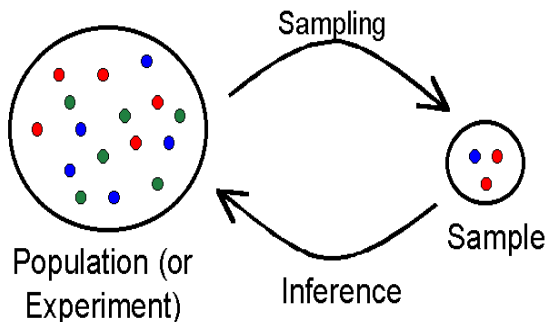
- 1 Sample size and sampling strategy
- 2 Variables measurement level
- 3 Hypotheses
- 4 Full model (functional form, interactions and specification
- 5 Methods and techniques
- 6 Inclusion, exclusion, transformations and corrections
- 7 Source of data
- 8 Computational software's
- 9 Research design limitations

10 features of PAP

- 1 Sample size and sampling strategy
- 2 Variables measurement level
- 3 Hypotheses
- 4 Full model (functional form, interactions and specification
- 5 Methods and techniques
- 6 Inclusion, exclusion, transformations and corrections
- 7 Source of data
- 8 Computational software's
- 9 Research design limitations
- 10 Sharing the document

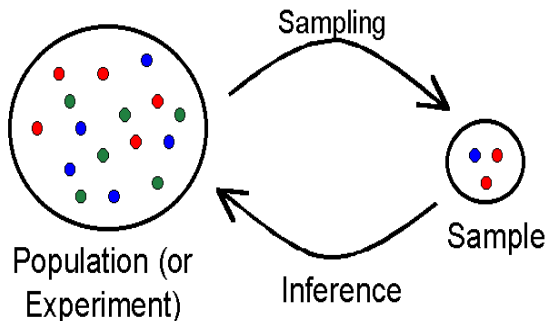
1. Sample size and sampling strategy

- 1 Define the unit of analysis (individuals, firms, cities, states, countries, etc.)



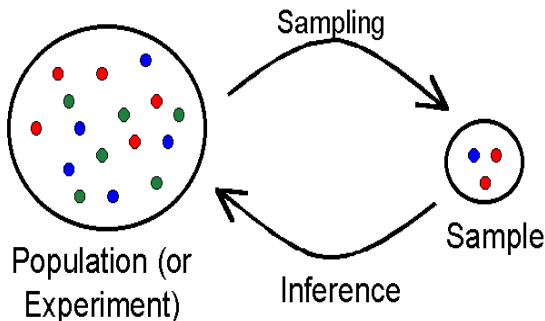
1. Sample size and sampling strategy

- 1 Define the unit of analysis (individuals, firms, cities, states, countries, etc.)
- 2 Sample size (margin of error, confidence intervals and sampling strategy)



1. Sample size and sampling strategy

- 1 Define the unit of analysis (individuals, firms, cities, states, countries, etc.)
- 2 Sample size (margin of error, confidence intervals and sampling strategy)
- 3 Logistics costs, time frame, data gathering problems



2. Variables Measurement Level

- 1 Measurement level of all variables (nominal, ordinal, discrete or continuous)

Example

“the only way to understand and evaluate an empirical analysis fully is to know the exact process by which the data were generated and the analysis produced” (KING 1995: 444).



2. Variables Measurement Level

- 1 Measurement level of all variables (nominal, ordinal, discrete or continuous)
- 2 Measurements strategies (unidimensional index, multidimensional, data reduction, aggregate measures)

Example

“the only way to understand and evaluate an empirical analysis fully is to know the exact process by which the data were generated and the analysis produced” (KING 1995: 444).

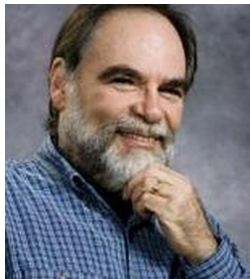


3. Hypotheses

- 1 Present the signal (+ or -)

Fact

in causal analysis, a hypothesis is a conjecture about the relationship between one or more independent variables and a dependent variable" (COLLIER, MAHONEY and SEAWRIGHT, 2004: 289)

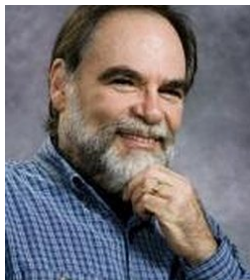


3. Hypotheses

- 1 Present the signal (+ or -)
- 2 Present the magnitude of the effect (big, moderate, small)

Fact

in causal analysis, a hypothesis is a conjecture about the relationship between one or more independent variables and a dependent variable" (COLLIER, MAHONEY and SEAWRIGHT, 2004: 289)



4. Model specification

- Functional form, interactions and specifications

Example

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

4. Model specification

- Functional form, interactions and specifications

Example

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

- Y equals student grade, X_1 represents teachers quality, X_2 identifies school location (urban or rural) and X_3 equals socioeconomic status

4. Model specification

- Functional form, interactions and specifications

Example

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

- Y equals student grade, X_1 represents teachers quality, X_2 identifies school location (urban or rural) and X_3 equals socioeconomic status
- The model is linear, assumes additivity and has only one specification

5. Data Management



- 1 Inclusion, exclusion, transformations, imputations

5. Data Management



- 1 Inclusion, exclusion, transformations, imputations
- 2 Outliers?

5. Data Management



- 1 Inclusion, exclusion, transformations, imputations
- 2 Outliers?
- 3 Identification techniques

5. Data Management



- 1 Inclusion, exclusion, transformations, imputations
- 2 Outliers?
- 3 Identification techniques
- 4 Accommodation methods

6. Methods

Example

“all data and analyses should, insofar as possible, be replicable (...) only by reporting the study in sufficient detail so that it can be replicated is it possible to evaluate the procedures followed and methods used” (KING, KEOHANE and VERBA, 1994: 26).

- 1 Describe all methodological choices

6. Methods

Example

“all data and analyses should, insofar as possible, be replicable (...) only by reporting the study in sufficient detail so that it can be replicated is it possible to evaluate the procedures followed and methods used” (KING, KEOHANE and VERBA, 1994: 26).

- 1 Describe all methodological choices
- 2 Report scripts

6. Methods

Example

“all data and analyses should, insofar as possible, be replicable (...) only by reporting the study in sufficient detail so that it can be replicated is it possible to evaluate the procedures followed and methods used” (KING, KEOHANE and VERBA, 1994: 26).

- 1 Describe all methodological choices
- 2 Report scripts
- 3 Share data

7. Data source



- Bibliography

7. Data source



- Bibliography
- Links

7. Data source



- Bibliography
- Links
- Dataverse (Harvard)

7. Data source

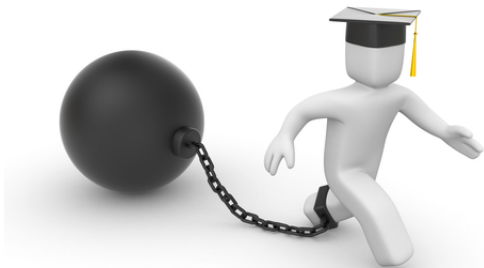


- Bibliography
- Links
- Dataverse (Harvard)
- CIS (USP)

8. Software



9. Research design limitations



10. Make your PAP public

- Dataverse

10. Make your PAP public

- Dataverse
- Academia.edu

10. Make your PAP public

- Dataverse
- Academia.edu
- ResearchGate

10. Make your PAP public

- Dataverse
- Academia.edu
- ResearchGate
- Personal website

10. Make your PAP public

- Dataverse
- Academia.edu
- ResearchGate
- Personal website
- OSF

10. Make your PAP public

- Dataverse
- Academia.edu
- ResearchGate
- Personal website
- OSF
- Facebook

10. Make your PAP public

- Dataverse
- Academia.edu
- ResearchGate
- Personal website
- OSF
- Facebook
- CIS

How to learn more about transparent research?

- Berkeley Initiative for Transparency in the Social Sciences (BITSS)

How to learn more about transparent research?

- Berkeley Initiative for Transparency in the Social Sciences (BITSS)
- Open Science Framework (OSF)

How to learn more about transparent research?

- Berkeley Initiative for Transparency in the Social Sciences (BITSS)
- Open Science Framework (OSF)
- Github - [here](#) and [here](#)

How to learn more about transparent research?

- Berkeley Initiative for Transparency in the Social Sciences (BITSS)
- Open Science Framework (OSF)
- Github - here and here
- SWIRL

How to learn more about transparent research?

- Berkeley Initiative for Transparency in the Social Sciences (BITSS)
- Open Science Framework (OSF)
- Github - [here](#) and [here](#)
- SWIRL
- Latex

How to learn more about transparent research?

- Berkeley Initiative for Transparency in the Social Sciences (BITSS)
- Open Science Framework (OSF)
- Github - here and here
- SWIRL
- Latex
- R

How to learn more about transparent research?

- Berkeley Initiative for Transparency in the Social Sciences (BITSS)
- Open Science Framework (OSF)
- Github - here and here
- SWIRL
- Latex
- R
- Reproducible research

How to learn more about transparent research?

- Berkeley Initiative for Transparency in the Social Sciences (BITSS)
- Open Science Framework (OSF)
- Github - here and here
- SWIRL
- Latex
- R
- Reproducible research
- Replication