



COURSE..... Policy Analytics with R - From Data Preparation to Impact Evaluation
SEMESTER/YEAR.... 2º/2026
PROGRAM..... Winter School
CLASS-HOURS..... ☐ 30 hours
PROFESSOR.....
LANGUAGE..... ENGLISH

COURSE OBJECTIVES

Provide participants with advanced skills in the use of statistical, econometric, and machine learning methods applied to the monitoring and evaluation of public policies. The aim is to consolidate essential theoretical knowledge, develop the ability to select analytical techniques appropriate to different application contexts, and foster skills in the R environment, while stimulating analytical and critical thinking in the interpretation of results.

LEARNING GOALS

The course learning goals are presented in the table below, showing how they contribute to the learning goals related to methods, for the stricto sensu graduate programs at FGV-EAESP.

Program learning goals	Course learning goals	Level of contribution
Scientific method	Application of the scientific method to policy evaluation, with emphasis on hypothesis formulation, causal framework, and rigorous interpretation of results	● ● ●
Research project / procedures	Learn tools for empirical research design, data analysis, project development, and preparation of scientific outputs.	● ● ○
Qualitative research methods	Learn qualitative approaches as complementary to quantitative analysis, highlighting their role in interpretation and mixed-methods research	● ○ ○
Quantitative research methods	Foster the skills in advanced quantitative methods (statistical, econometric, and machine learning) for public policy evaluation	● ● ●
Development of academic papers	Develop the ability to prepare a scientific publication, structured as "short communication", reporting the application of one method to research data (element for the assessment of the course)	● ● ○
<u>Other course learning goals:</u> exposure to the most recent methodological approaches and applications, development of coding skills in R, and stimulation of analytical and critical thinking		

The full description of the learning goals of FGV-EAESP stricto sensu graduate programs can be found at <https://rebrand.ly/cursos-pos-eaesp>.

PREVIOUS KNOWLEDGE REQUIRED

The course is designed for researchers, senior technicians, data analysts, and public decision-makers who are interested in applying quantitative methodologies to the monitoring and evaluation of public policies, thereby informing and supporting decision-making processes. It is also suitable for undergraduate and graduate students in Public Administration, Public Policy, Social Policy, Political Science, Economics, Management, and related areas.

Basic knowledge of descriptive statistics (e.g., measures of central tendency, position, and dispersion), inferential statistics (e.g., point and interval estimation, hypothesis testing), and principles of sampling is recommended. No prior programming knowledge is required.



CONTENT/METHODOLOGY

Teaching methodologies

Learning methods will be used that involve students in the analysis and application of concepts in real contexts, resulting in the acquisition of essential theoretical and methodological skills. To this end, the modules will be developed using data from the scientific publications to provide a reference for comparing the results obtained during the application of the methodologies and for contextualizing their interpretation.

The content will be addressed and applied using the open source software RStudio, and with a coding assistant supported by Artificial Intelligence, to introduce greater dynamism into learning and to facilitate the application of the R language among participants who have no prior experience of programming.

Program content

Module 1 - Quantitative methods for monitoring and evaluating public policies.

Learning content and outcomes - Learn the principles and configurations of experimental, quasi-experimental, and observational studies. Understand the objectives, methodological approaches, and applications of the main methods, including the latest methodologies based on machine learning.

Reference publications for illustrating examples - Abadie & Cattaneo (2018). <https://doi.org/10.1146/annurev-economics-080217-053402> | Athey & Imbens (2019). <https://doi.org/10.1146/annurev-economics-080217-053433> | Cashin et al. (2025) <https://doi.org/10.1136/bmj-2025-087179>.

Module 2 - Textual analysis for document mapping and thematic characterization.

Learning content and outcomes - Develop skills in text data mining (including pre-processing techniques) and in the use of algorithms to identify and characterize latent themes. Learn to classify and relate documents based on the identified themes and their associated words.

Reference publication for the data source - Vikstrom et al. (2025). <https://doi.org/10.1016/j.enpol.2025.114746>.

Module 3 - Descriptive analysis and data visualization for public policy monitoring.

Learning content and outcomes - Learn descriptive statistical measures for data summarization, factorization techniques for index construction, and techniques for grouping observed units. Develop data visualization skills to facilitate the reporting of indicators and results.

Reference publications for the data sources - Jugl (2025). <https://doi.org/10.1093/jopart/muaf012> | Quarterly Census of Employment and Wages (QCEW) <https://www.bls.gov/cew/>.

Module 4 - Regression analysis for causal inference.

Learning content and outcomes - Learn regression techniques for causal inference, including adjustment for observed covariates. Introduce weighting methods using propensity scores estimated by inverse probability and doubly robust methods to reduce bias and improve the validity of estimates.

Reference publications for the data sources - Bodory et al. (2025). <https://CRAN.R-project.org/package=causalweight> | Sekhon (2021) <https://doi.org/10.1111/biomet.12444>.



Module 5 - Machine learning methods for causal inference.

Learning content and outcomes - Learn machine learning algorithms applied to causal inference, with an emphasis on regularisation, decision trees, and random forests. Analyse the heterogeneity of effects to capture different patterns of impact in subgroups of interest.

Reference publications for the data sources - Bodory et al. (2025). <https://CRAN.R-project.org/package=causalweight> | Sekhon (2021). <https://doi.org/article/b6d400df168b48619b4d707cb6b849af>.

Module 6 - Discontinuous regression method.

Learning content and outcomes - Learn how to apply the discontinuous regression method, both in deterministic (sharp) and probabilistic (fuzzy) formats, including bandwidth analysis and its implications for the robustness of estimates.

Reference publications for the data sources - Lundqvist et al. (2014). <https://doi.org/10.3886/E114847V1> | Nguyen & Rieger (2016). <https://doi.org/10.1016/j.worlddev.2016.10.013>.

Module 7 - Difference-in-differences method.

Learning content and outcomes - Learn the difference-in-differences method, with and without covariate adjustment. Introduce extensions that consider multiple intervention periods. Learn to assess the balance of groups concerning covariates of interest.

Reference publications for the data sources - Callaway & Sant'Anna (2021). <https://doi.org/10.1016/j.jeconom.2020.12.001> | Sant'Anna & Zhao (2020). <https://doi.org/10.1016/j.jeconom.2020.06.003>.

Module 8 - Synthetic control technique.

Learning content and outcomes - Learn the synthetic control technique as a method for evaluating public policies through the construction of counterfactual units. Assess its applicability, especially in single case studies or regional comparisons.

Reference publications for the data sources - Arkhangelsky et al. (2021). <https://doi.org/10.48550/arXiv.1812.09970>.

EVALUATION CRITERIA

The assessment will be based on a scientific article, in the form of a “short communication”, reporting the main conclusions obtained from the application of one of the methods learned during the course, using the participant's own research data (or data provided to participants by the lecturer). The scientific article should be delivered up to 10 weekdays upon the end of the course.



COURSE SCHEDULE (OPTIONAL)

Program learning goals	Length
Module 1 - Quantitative methods for monitoring and evaluating public policies	3h40m
Module 2 - Textual analysis for document mapping and thematic characterization	3h40m
Module 3 - Descriptive analysis and data visualization for public policy monitoring	3h40m
Module 4 - Regression analysis for causal inference	3h40m
Module 5 - Machine learning methods for causal inference	3h40m
Module 6 - Discontinuous regression method	3h40m
Module 7 - Difference-in-differences method	3h40m
Module 8 - Synthetic control technique	3h40m

BIBLIOGRAPHY

Békés, G., & Kézdi, G. (2021). Data analysis for business, economics, and policy. Cambridge University Press.

Chen, J. C., Rubin, E. A., & Cornwall, G. J. (2021). Data science for public policy. Springer.

Cunningham, S. (2021). Causal inference. Yale University Press.

Gertler, Paul J., Martinez, Sebastian, Premand, Patrick, Rawlings, Laura B., & Vermeersch, Christel M. J. (2016). *Impact Evaluation in Practice* (2nd ed.). World Bank Publications.

Huntington-Klein, N. (2025). The effect: An introduction to research design and causality. (2nd ed.). Chapman and Hall/CRC.

FURTHER INFORMATION (OPTIONAL)