

mediator: an R package for implementing causal mediation analyses

Jordan Creed and Travis Gerke

Background: Mediation analysis allows users to examine the relationship between an exposure and outcome through an intermediate variables, the mediator variable, which can be used to further characterize the relationship between the exposure and outcome. The counterfactual framework of mediation analysis provides an extension of the work by Baron and Kenny by allowing for exposure-mediator interactions and the decomposition of the total effect into direct and indirect effects.

Previously software for performing this type of analysis was made available as SAS and SPSS macros by Valeri and VanderWeele, while within R, the package `mediation` provides some of these functions but not all and uses a different set of terminology and framework.

We have created an R package `mediator` which provides point estimates and confidence intervals for the controlled direct effect (CDE), as well as the natural direct and indirect effects (NDE and NIE), total effect (TE) and proportion mediated (PM) under multiple conditions and allows, but does not require, interaction between the exposure and mediator variables.

Usage: The package was developed using R v 3.6.1 and is available on github and can be installed and brought into the environment using the following commands `devtools::install_github("gerkelab/mediator")` ; `library(mediator)`. The `mediator` package allows for binary and continuous exposures, mediators and outcomes as well as survival outcomes.

At a minimum users need to provide the data set for the analysis (`data =`), a model for the outcome and mediator (`out.model =` and `med.model =`), and specify what type of models are being used (`out.reg =` and `med.reg =`). Users also need to specify which variables are the treatment (`treat =`) and mediator (`mediator =`). The package assumes proper modeling on the part of the user. Additional options include setting the exposure level (`a =`), the compared exposure level (`a_star =`), the level of the mediator (`m =`) and the number of bootstrap replications for calculating the confidence interval (`boot_rep =`). The level of the mediator is used for calculating the CDE and thus there are as many potential values for the CDE as there are levels of the mediator. By default the function calculates confidence intervals using the Delta method, but by changing the number of bootstrap replicates the user alters how the confidence interval is calculated automatically.

The function returns a tibble with the CDE, NDE, NIE, TE and their 95% confidence intervals along with the PM. Covariates for the returned effects are either the mean value (continuous) or the most common value (categorical).

Example: Using the Surveillance, Epidemiology, and End Results (SEER) database, we examine the effect of race on insurance status and whether it is mediated through Yost Index, which is a census-track level composite measure of socio-economic status indicators. When setting the mediator level (Yost index) equal to the median value in Caucasian and African American men, being African American was protective against being insured and roughly the same regardless of what Yost Index was set at (OR = 0.44; 95% CI 0.42, 0.47 and OR = 0.45; 95% CI 0.43, 0.47, respectively). These results are compatible with the low proportion mediated of 0.06.

Conclusion: The `mediator` R package provides an efficient mechanism for conducting casual mediation analyses and a useful tool in reproducible epidemiologic research.