Package ‘domir’

March 29, 2021

Title Tools to Support Relative Importance Analysis
Version 0.0.1
Date 2021-3-23

Description Provides tools that support relative importance analysis focusing on dominance analysis. Dominance analysis is a methodology for determining the relative importance of predictors/features/independent variables (Azen, R., & Budescu, D. V. (2003) <doi:10.1037/1082-989X.8.2.129>; Groemping, U. (2007) <doi:10.1198/000313007X188252>); as well as parameter estimates (Luchman, J. N, Lei, X., & Kaplan, S. (2020) <doi:10.47263/JASEM.4(2)02>). These tools are intended to extend relative importance analysis to, effectively, any statistical or machine learning function as defined or desired by the user-especially those where the user wants to use custom importance/fit metric or modeling function.

Imports methods
Suggests dominanceanalysis, relaimpo, yhat
License GPL (>= 2)

URL https://github.com/jluchman/domir
Encoding UTF-8

R topics documented:

  domir-package .......................................................... 2
domin ................................................................. 3
print.domin ........................................................... 5
Description

The domir package provides a set of flexible wrapper and helper functions for conducting relative importance analysis with a focus on dominance analysis. The intention of this package is to provide tools that allow relative importance analysis across a wide variety of practical data analytic situations.

Details

Relative importance (RI) analysis is a methodology focused on comparing independent variables/features/predictors as well as parameter estimates to one another in terms of how they predict some dependent variable/response/outcome. The methodology implemented in this initial version of domir is dominance analysis (DA).

As a RI method, DA determines the relative importance of independent variables (IVs) or parameter estimates (PEs) in an estimation model based on contribution to an overall model fit statistic/metric (see Budescu, 1993; Groemping, 2007 for a discussions). DA is an ensemble method in which importance determinations about IV/PEs are made by aggregating results across multiple models, though the method usually requires the ensemble contain each possible combination of the IV/PEs in the full model that has been selected using model selection methods.

The all possible combinations ensemble with p IV/PEs in the full model results in $2^p$ models estimated. That is, each combination of p IV/PEs alternating between included versus excluded (i.e., the base of 2 to the p exponent number of IV/PEs).

Currently, the only tool implemented in domir is a DA method domin. domin is a flexible wrapper function that can be used with many modeling functions that accept a formula that follow the standard \texttt{response ~ terms} format. The format used by domin can be extended to other functions focused on IV-based RI can be accommodated with a additional wrapper functions based on the formula it creates and submits to modeling functions.

Author(s)

Joseph Luchman <jluchman_at_gmail_com>

References

Dominance analysis supporting formula-based functions

Description

Computes dominance statistics for predictive modeling functions that accept a "standard" formula.

Usage

domin(
  formula_overall,
  reg,
  fitstat,
  sets = NULL,
  all = NULL,
  complete = TRUE,
  ...
)

Arguments

formula_overall
  An object of class formula or that can be coerced to class formula for use in the function in reg. The formula object must have the form response ~ terms with the terms separated by +.

reg
  The function implementing the predictive (or "reg"ression) model called. Uses do.call and accepts any function call do.call would accept.

fitstat
  List of specifications to call a fit statistic extracting function (see details). Like reg, uses do.call and accepts any function call do.call would accept.

sets
  An optional list containing vectors of variable or factor names. Each vector in the list is used as a set and always included together in the formula.

all
  An optional vector of variable or factor names to be built into the formula and included in all model subsets.

complete
  Logical. If FALSE then complete dominance matrix is not computed.

...
  Additional arguments passed to the function in reg.
**Details**

Dominance analysis focuses on computing the contribution of independent variables/IVs or terms to a predictive model's fit to the data. domin automates the process whereby combinations of IVs are created and concatenated in a formula to be submitted to the model in reg. domin creates all the concatenated IVs from the entries on the right hand side of formula_overall and the entries in sets.

Each entry in the right of formula_overall is processed and each individual entry is included as a separate IV. formula_overall must contain the dependent variable/DV or response on the left hand side and any individual IVs separated by +. domin applies only the formula processing that is available in the stats package.

The elements of the list entries in sets are each considered a separate IV and must be submitted as a list. Each entry in sets must be a vector of IVs. Individual vector elements within a single set are concatenated using + automatically. It is possible to use a domin with only sets (i.e., no IVs in formula_overall; see examples below).

The IV's in all must also be submitted as a vector, are concatenated with + automatically, and are also included in the model. These "all subsets" IVs are removed from the fit statistic and all subsequent dominance statistics.

The entry to fitstat must be list and follow a specific structure: (fit_function, element_name, ...)

- **fit_function** First element and function to be applied to reg
- **element_name** Second element and name of the element from the object returned by fit_function
- ... Subsequent elements and are additional arguments passed to fit_function

**Value**

Returns an object of class "domin". An object of class "domin" is a list composed of the following elements:

- **General_Dominance** Vector of general dominance statistics.
- **Standardized** Vector of general dominance statistics normalized to be out of 100.
- **Ranks** Vector of ranks applied to the general dominance statistics.
- **Conditional_Dominance** Matrix of conditional dominance statistics.
- **Complete_Dominance** Matrix of complete dominance designations.
- **Fit_Statistic_Overall** Value of fit statistic across all IVs.
- **Fit_Statistic_All_Sets** Value of fit statistic associated with IVs in all.
- **Call** The matched call.
- **Subset_Details** List containing full model and descriptions of IVs in model by source.

**Examples**

```r
## Basic linear model with r-square
domin(mpg ~ am + vs + cyl, "lm", list("summary", "r.squared"), data=mtcars)

## Including sets
```
print.domin

```r
# Multivariate linear model with custom multivariate r-square function and all subsets variable
Rxy <- function(obj, names, data) {
  return(list("r2" = cancor(predict(obj),
        as.data.frame(mget(names, as.environment(data))))[["cor"]][1]^2))
}

# Sets only
domin(mpg ~ 1, "lm", list("summary", "r.squared"),
      data = mtcars, sets = list(c("cars", "gear"), c("disp", "wt")))
```

---

**Description**

Reports basic results from `domin` class object.

**Usage**

```r
## S3 method for class 'domin'
print(x, ...)```

**Arguments**

- `x`: an object of class "domin".
- `...`: further arguments passed to or from other methods.

**Details**

The print method for class `domin` objects reports out the following results:

- Fit statistic for the full model as well as the fit statistic for the all subsets model if any entries in all.
- Matrix describing general dominance statistics, standardized general dominance statistics, and the ranking of the general dominance statistics.
- Matrix describing the conditional dominance statistics.
- If conditional is TRUE, matrix describing the complete dominance statistics.

**Value**

No returned value. This method is called for compact display of results in the console.
Index

* multivariate
  domin, 3
* print
  print.domin, 5
* utilities
  domin, 3
class, 4
do.call, 3
domin, 3
domir (domir-package), 2
domir-package, 2

formula, 3

print.domin, 5

vector, 3