Package: APCI (via r-universe)

September 2, 2024

Type Package

Title A New Age-Period-Cohort Model for Describing and Investigating Inter-Cohort Differences and Life Course Dynamics

Version 1.0.8

Maintainer Jiahui Xu <jpx5053@psu.edu>

Depends R (>= 3.6.0)

Description It implemented Age-Period-Interaction Model (APC-I Model) proposed in the paper of Liying Luo and James S. Hodges in 2019. A new age-period-cohort model for describing and investigating inter-cohort differences and life course dynamics.

Imports survey, magrittr, dplyr, ggplot2, data.table, ggpubr, stringr, gee

License GPL-2

Encoding UTF-8

LazyData true

Repository https://jiahui1902.r-universe.dev

RemoteUrl https://github.com/jiahui1902/apci

RemoteRef HEAD

RemoteSha 8df0788a3a1d73044ae99ede23d8df01006f689d

Contents

`ageperiod_group` *Construct a cohort index matrix for any number of age and period groups*

Description

This function returns a cohort index matrix for any number of age and period groups. The cohort index matrix will then be used to extract age-period interaction effects contained in each cohort.

Usage

```
ageperiod_group(age_range, period_range,
age_interval, period_interval,
age_group = NULL, period_group = NULL)
```

Arguments

`age_range`, `period_range`
: Numeric vector indicating the actual age and period range (e.g., 10 to 59 years old from 2000 to 2019).

`age_interval`, `period_interval`, `age_group`, `period_group`
: Numeric values or character vectors indicating how age and period are grouped. `age_interval` and `period_interval` are numbers indicating the width of age and period groups respectively. `age_group` and `period_group` are character vectors explicitly listing all potential age and period groups. Either `age_interval`(`period_interval`) or `age_group` (`period_group`) have to be defined when `unequal_interval` is `TRUE`.

Value

It returns a matrix respresenting the relationship among age, period, and cohort groups under the current setting.

Examples

```
## age and period groups have equal width
ageperiod_group(age_range = 10:59, period_range = 2000:2019,
                age_interval = 5, period_interval = 5)
ageperiod_group(age_range = 10:59, period_range = 2000:2019,
                age_group = c("10-14","15-19","20-24","25-29",
                              "30-34","35-39","40-44","45-49",
                              "50-54","55-59"),
                period_group = c("2000-2004","2005-2009","2010-2014","2015-2019"))

## age and period groups have unequal width
ageperiod_group(age_range = 10:59, period_range = 2000:2019,
                age_interval = 10, period_interval = 5)
ageperiod_group(age_range = 10:59, period_range = 2000:2019,
                age_group = c("10-19","20-29","30-39","40-49","50-59"),
                period_group = c("2000-2004","2005-2009","2010-2014","2015-2019"))
```

`apci` *Run apci model*

Description

run APC-I model

Arguments

`outcome`	An object of class character containing the name of the outcome variable. The outcome variable can be continuous, categorical, or count.
`age`	An object of class character representing the age group index taking on a small number of distinct values in the data. Usually, the vector should be converted to a factor (or the terms of "category" and "enumerated type").
`period`	An object of class character, similar to the argument of age, representing the time period index in the data.
`cohort`	An optional object of class character representing cohort membership index in the data. Usually, the cohort index can be generated from the age group index and time period index in the data because of the intrinsic relationship among these three time-related indices.
`weight`	An optional vector of sample weights to be used in the model fitting process. If non-NULL, the weights will be used in the first step to estimate the model. Observations with negative weights will be automatically dropped in modeling.
`covariate`	An optional vector of characters, representing the name(s) of the user-specified covariate(s) to be used in the model. If the variable(s) are not found in data, there will be an error message reminding the users to check the data again.
`data`	A data frame containing the outcome variable, age group indicator, period group indicator, and covariates to be used in the model. If the variable(s) are not found in data, there will be an error message reminding the users to check the input data again.

`family` Used to specify the statistical distribution of the error term and link function to be used in the model. Usually, it is a character string naming a family function. For example, family can be "binomial", "multinomial"", or "gaussian". Users could also check R package glm for more details of family functions.

`dev.test` Logical, specifying if the global F test (step 1) should be implemented before running the APC-I model. If TRUE, apci will first run the global F test and report the test results; otherwise, apci will skip this step and return NULL. The default setting is TRUE. But users should be careful that the algorithm will not automatically stop even if there is no significant cohort average deviation.

`print` Logical, specifying if the intermediate results should be displayed on the screen in running the model. The default setting is TRUE in order to show the results explicitly although it can be too clumpy when the intermediate results are shown on the screen.

`gee` logical, indicating if the data is cross-sectional data or longitudinal/panel data. If `TRUE`, the generalized estimating equation will be used to correct the standard error estimates. The default is `FALSE`, indicating that the data are cross-sectional.

`id` A vector of character, specifying the cluster index in longitudinal data. It is required when `gee` is `TRUE`. The length of the vector should be the same as the number of observations.

`corstr` a character string, specifying a possible correlation structure in the error terms when `gee` is `TRUE`. The following are allowed: `independence`, `fixed`, `stat\_M\_dep`, `non\_stat\_M\_dep`, `exchangeable`, `AR-M` and `unstructured`. The default value is `exchangeable`.

`unequal_interval` Logical, indicating if age and period groups are of the same width. The default is set as `TRUE`.

`age_range`, `period_range` Numeric vector indicating the actual age and period range (e.g., 10 to 59 years old from 2000 to 2019).

`age_interval`, `period_interval`, `age_group`, `period_group` Numeric values or character vectors indicating how age and period are grouped. `age_interval` and `period_interval` are numbers indicating the width of age and period groups respectively. `age_group` and `period_group` are character vectors explicitly listing all potential age and period groups. Either `age_interval`(`period_interval`) or `age_group` (`period_group`) have to be defined when `unequal_interval` is `TRUE`.

`...`

Value

`model` A summary of the fitted generalized linear regression. It displays the coefficients, standard errors, etc.

`dev_global` The results of the global F test. It shows that if the interaction terms are significant as a component of the generalized linear regression model.

`intercept` The overall intercept.

`age_effect` A vector, representing the estimated age effect for each age group.

period_effect	A vector, representing the estimated period effect for each time period.
cohort_average	A vector, representing the cohort average effects for comparing inter-cohort differences.
cohort_slope	A vector, representing intra-cohort life-course changes.

Examples

```
library("APCI")
## load data
test_data <- APCI::women9017
test_data$acc <- as.factor(test_data$acc)
test_data$pcc <- as.factor(test_data$pcc)
## run APCI model
APC_I <- apci(outcome = "inlfc",
              age = "acc",
              period = "pcc",
              cohort = "ccc",
              weight = "wt",
              data = test_data,dev.test=FALSE,
              family = "gaussian")

## check model results
summary(APC_I)

APC_I$model
APC_I$dev_global
APC_I$dev_local
APC_I$intercept
APC_I$age_effect
APC_I$period_effect
APC_I$cohort_average
APC_I$cohort_slope
```

---

apci.bar *Visualization for the APC-I model results*

Description

Visualize the APC-I model results in a simple bar plot.

Usage

```
apci.bar(model, age, period, outcome_var,
cohort_label = NULL, ...)
```

Arguments

model	A list, inheriting the corresponding results generated by function apci.
age	A vector, representing the age group index taking on a small number of distinct values in the data. Usually, the vector should be converted to a factor (or the terms of "category" and "enumerated type").
period	An object of class character, similar to the argument of age, representing the time period index in the data.
outcome_var	An object of class character representing the name of the outcome variable used in APC-I model. The outcome variable itself can be numerical and categorical.
cohort_label	A vector, representing the labels of cohort groups in the x asix.
...	

Examples

```
library("APCI")
## load data
test_data <- APCI::women9017
test_data$acc <- as.factor(test_data$acc)
test_data$pcc <- as.factor(test_data$pcc)

## run APCI model
APC_I <- apci(outcome = "inlfc",
              age = "acc",
              period = "pcc",
              cohort = "ccc",
              weight = "wt",
              data = test_data,dev.test=FALSE,
              family = "gaussian")

## plot the bar plot
apci.bar(model = APC_I, age = "acc",period = "pcc")
```

---

apci.plot *Visualization for data exploration or model results*

Description

Visualize the outcome or APC-I model results in a simple plot.

Usage

```
apci.plot(model, age, period, outcome_var,
type = "model", quantile = NULL, ...)
```

Arguments

model	A list, inheriting the corresponding results generated by function apci.
outcome_var	An object of class character representing the name of the outcome variable used in APC-I model. The outcome variable itself can be numerical and categorical.
age	An object of class character representing the age group index taking on a small number of distinct values in the data. Usually, the vector should be converted to a factor (or the terms of "category" and "enumerated type").
period	An object of class character, similar to the argument of age, representing the time period index in the data.
type	Character, "explore" or "model". If type is "explore", plots for age and period raw scores will be generated. If type is "model", model results will be plotted. The default setting is "model".
quantile	A number between 0 and 1, representing the percentiles to be used in visualizing the data or model. If NULL, the original magnitude will be used.
...	

---

apci.plot.heatmap *Visualize APC-I model results*

Description

Visualize cohort effects from the APC-I model results using a heatmap

Usage

```
apci.plot.heatmap(model, age, period, color_map = NULL,
color_scale = NULL, quantile = NULL, ...)
```

Arguments

model	A list, inheriting the corresponding results generated by function apci.
age	A vector, representing the age group index taking on a small number of distinct values in the data. Usually, the vector should be converted to a factor (or the terms of "category" and "enumerated type").
period	An object of class character, similar to the argument of age, representing the time period index in the data.
color_map	A vector, representing the color palettes to be used in the figure. The default setting is greys if color_map is NULL. Alternations, for example, can be c("blue", "yellow"), blues, etc.
color_scale	A vector including two numbers indicating the limit of the values to be plotted. The first number is the minimum value to be visualized and the second is the maximum value to be visualized. If NULL, the algorithm will automatically select the limits from the data (estimation results) to set up the scale.
quantile	A number between 0 and 1, representing the percentiles to be used in visualizing the data or model. If NULL, the original magnitude will be used.
...	

Examples

```
library("APCI")
## load data
test_data <- APCI::women9017
test_data$acc <- as.factor(test_data$acc)
test_data$pcc <- as.factor(test_data$pcc)

## run APCI model
APC_I <- apci(outcome = "inlfc",
              age = "acc",
              period = "pcc",
              cohort = "ccc",
              weight = "wt",
              data = test_data,dev.test=FALSE,
              family = "gaussian")

## plot heatmap
apci.plot.heatmap(model = APC_I, age = "acc",period = 'pcc',
                  color_map = c('blue','yellow'))
```

---

apci.plot.hexagram *Visualize APC-I model results*

Description

Visualize cohort effects from the APC-I model results using a hexagram

Usage

```
apci.plot.hexagram(model, age, period, first_age,
first_period, interval, first_age_isoline = NULL,
first_period_isoline = NULL, isoline_interval = NULL,
color_scale = NULL, color_map = NULL, line_width = 0.5,
line_color = "grey", label_size = 0.5,
label_color = "black", scale_units = "Quintile",
wrap_cohort_labels = TRUE, quantile = NULL)
```

Arguments

model	A list, inheriting the corresponding results generated by function apci.
age	An object of class character representing the age group index taking on a small number of distinct values in the data. Usually, the vector should be converted to a factor (or the terms of "category" and "enumerated type").
period	An object of class character, similar to the argument of age, representing the time period index in the data.

color_scale A vector including two numbers indicating the limit of the values to be plotted. The first number is the minimum value to be visualized and the second is the maximum value to be visualized. If NULL, the algorithm will automatically select the limits from the data (estimation results) to set up the scale.

color_map A vector, representing the color palettes to be used in the figure. The default setting is greys if color_map is NULL. Alternations, for example, can be c("blue", "yellow"), blues, etc.

first_age

first_period

interval

first_age_isoline

first_period_isoline

isoline_interval

line_width

line_color

label_size

label_color

scale_units

wrap_cohort_labels

quantile

---

apci.plot.raw *data exploration: visualize age, period, and cohort patterns in the outcome before modeling.*

Description

visualize age, period, and cohort patterns before modeling.

Usage

```
apci.plot.raw(data, outcome_var, age, period, ...)
```

Arguments

data A data frame containing the outcome variable, age group indicator, period group indicator, and covariates to be used in the model. If the variable(s) are not found in data, there will be an error message reminding the users to check the input data again.

outcome_var	An object of class character containing the name of the outcome variable. The outcome variable can be continuous, categorical, or count.
age	An object of class character representing the age group index taking on a small number of distinct values in the data. Usually, the vector should be converted to a factor (or the terms of "category" and "enumerated type").
period	An object of class character, similar to the argument of age, representing the time period index in the data.
...	

Examples

```
# plot the raw scores
apci.plot.raw(data = simulation, outcome_var = "y",
              age = "age", period = "period")
```

---

cohortdeviation *compute inter-cohort avarage deviations and intra-cohort life-course dynamics*

Description

Compute inter- and intra-cohort deviations.

Usage

```
cohortdeviation(A,
  P,
  C,
  model = temp6,
  weight = "wt",
  covariate,
  gee=FALSE,
  unequal_interval = FALSE,
  age_range = NULL,
  period_range = NULL,
  age_interval = NULL,
  period_interval = NULL,
  age_group = NULL,
  period_group = NULL,
  ...)
```

Arguments

A, P, C	The numbers of age groups, period groups, and cohort groups are defined separately.
model	A generalized linear model generated from the internal function temp_model

`weight` An optional vector of sample weights to be used in the model fitting process. If non-NULL, the weights will be used in the first step to estimate the model. Observations with negative weights will be automatically dropped in modeling.

`covariate` An optional vector of characters, representing the name(s) of the user-specified covariate(s) to be used in the model. If the variable(s) are not found in data, there will be an error message reminding the users to check the data again.

`gee` logical, indicating if the data is cross-sectional data or longitudinal/panel data. If `TRUE`, the generalized estimating equation will be used to correct the standard error estimates. The default is `FALSE`, indicating that the data are cross-sectional.

`unequal_interval` Logical, indicating if age and period groups are of the same width. The default is set as `TRUE`.

`age_range`, `period_range` Numeric vector indicating the actual age and period range (e.g., 10 to 59 years old from 2000 to 2019).

`age_interval`, `period_interval`, `age_group`, `period_group` Numeric values or character vectors indicating how age and period are grouped. `age_interval` and `period_interval` are numbers indicating the width of age and period groups respectively. `age_group` and `period_group` are character vectors explicitly listing all potential age and period groups. Either `age_interval`(`period_interval`) or `age_group` (`period_group`) have to be defined when `unequal_interval` is `TRUE`.

`...`

`compute_xcoordinate` *calculate x coordinate values*

Description

Calculate x coordinate values for the hexagram. This is an intermediate function.

Usage

```
compute_xcoordinate(p)
```

Arguments

`p`

compute_ycoordinate *calculate y coordinate values*

Description

Calculate y coordinate values for the hexagram. This is an intermediate function.

Usage

```
compute_ycoordinate(p, a)
```

Arguments

p

a

cpsmen *Labor force participation data for men from 1990 to 1979 in CPS*

Description

the dataset for men

Usage

```
data("cpsmen")
```

Format

A data frame with 10000 observations on the following 7 variables.

`asecwt` weight

`year` a factor indicating period groups with levels `1 2 3 4 5 6`

`age` a factor indicating age groups with levels `1 2 3 4 5 6 7 8 9`

`labforce` labor Force participation rate

`educ` education level

`educr` education level

`educc` education level

cpswomen — *Women's labor force participation data from the 1990 to 2019 Current Population Survey (CPS)*

Description

the dataset for women's labor force participation from the 1990 through 2019 CPS.

Usage

```
data("cpswomen")
```

Format

A data frame with 1,0000 observations and the following 7 variables.

`asecwt` weight

`year` a factor indicating period groups with levels `1 2 3 4 5 6`

`age` a factor indicating age groups with levels `1 2 3 4 5 6 7 8 9`

`labforce` labor Force participation rate

`educ` education level

`educr` education level

`educc` education level

maineffect — *estimate age main effects and period main effects*

Description

estimate age and period main effects from the APCI model

Usage

```
maineffect(A, P, C, model = temp6, data, gee=FALSE,
...)
```

Arguments

A, P, C	The numbers of age groups, period groups, and cohort groups separately.
model	A generalized linear regression model generated from the internal function temp_model
data	A data frame containing the outcome variable, age group indicator, period group indicator, and covariates to be used in the model. If the variable(s) are not found in data, there will be an error message reminding the users to check the input data again.
gee	logical, indicating if the data is cross-sectional data or longitudinal/panel data. If TRUE, the generalized estimating equation will be used to correct the standard error estimates. The default is FALSE, indicating that the data are cross-sectional.
...	

---

simulation *Simulated Dataset*

Description

A simulated dataset for APC-I analysis.

Usage

```
data("simulation")
```

Format

A data frame with 1,0000 observations and the following 3 variables.

`y` a numeric

`age` a numeric

`period` a numeric

---

temp_model *Estimate APC-I model*

Description

Step 1 of the APCI model: estimate a generalized linear model.

Usage

```
temp_model(data,
           outcome = "inlfc",
           age = "acc",
           period = "pcc",
           cohort = NULL,
           weight = NULL,
           covariate = NULL,
           family = "quasibinomial",
           gee = FALSE,
           id = NULL,
           corstr = "exchangeable",
           ...)
```

Arguments

data	A data frame containing the outcome variable, age group indicator, period group indicator, and covariates to be used in the model. If the variable(s) are not found in data, there will be an error message reminding the users to check the input data again.
outcome	An object of class character containing the name of the outcome variable. The outcome variable can be continuous, categorical, or count.
age	An object of class character representing the age group index taking on a small number of distinct values in the data. Usually, the vector should be converted to a factor (or the terms of "category" and "enumerated type").
period	An object of class character, similar to the argument of age, representing the time period index in the data.
cohort	An optional object of class character representing cohort membership index in the data. Usually, the cohort index can be generated from the age group index and time period index in the data because of the intrinsic relationship among these three time-related indices.
weight	An optional vector of sample weights to be used in the model fitting process. If non-NULL, the weights will be used in the first step to estimate the model. Observations with negative weights will be automatically dropped in modeling.
covariate	An optional vector of characters, representing the name(s) of the user-specified covariate(s) to be used in the model. If the variable(s) are not found in data, there will be an error message reminding the users to check the data again.
family	Used to specify the statistical distribution of the error term and link function to be used in the model. Usually, it is a character string naming a family function. For example, family can be "binomial", "multinomial"", or "gaussian". Users could also check R package glm for more details of family functions.
gee	logical, indicating if the data is cross-sectional data or longitudinal/panel data. If TRUE, the generalized estimating equation will be used to correct the standard error estimates. The default is FALSE, indicating that the data are cross-sectional.

id	A vector of character, specifying the cluster index in longitudinal data. It is required when gee is TRUE. The length of the vector should be the same as the number of observations.
corstr	a character string, specifying a possible correlation structure in the error terms when gee is TRUE. The following are allowed: independence, fixed, stat_M_dep, non_stat_M_dep, exchangeable, AR-M and unstructured. The default value is exchangeable.
...	

---

tests *local and global F test*

Description

implement local and global F test for APCI model

Usage

```
tests(model, age = "acc", period = "pcc",
cohort = "ccc", A, P, C, data, weight = "wt",
family, outcome, ...)
```

Arguments

model	A generalized linear regression model generated from the internal function temp_model
age	An object of class character representing the age group index taking on a small number of distinct values in the data. Usually, the vector should be converted to a factor (or the terms of "category" and "enumerated type").
period	An object of class character, similar to the argument of age, representing the time period index in the data.
cohort	An optional object of class character representing cohort membership index in the data. Usually, the cohort index can be generated from the age group index and time period index in the data because of the intrinsic relationship among these three time-related indices.
A, P, C	The numbers of age groups, period groups, and cohort groups separately.
data	A data frame containing the outcome variable, age group indicator, period group indicator, and covariates to be used in the model. If the variable(s) are not found in data, there will be an error message reminding the users to check the input data again.
weight	An optional vector of sample weights to be used in the model fitting process. If non-NULL, the weights will be used in the first step to estimate the model. Observations with negative weights will be automatically dropped in modeling.
family	Used to specify the statistical distribution of the error term and link function to be used in the model. Usually, it is a character string naming a family function. For example, family can be "binomial", "multinomial"", or "gaussian". Users could also check R package glm for more details of family functions.

outcome An object of class character containing the name of the outcome variable. The outcome variable can be continuous, categorical, or count.

...

---

women9017 *women9017*

Description

A sample dataset

Usage

```
women9017
```

Format

A data frame with 1,000 observations and 23 variables.

ac a numeric vector

acc a numeric vector

age a numeric vector

cc a numeric vector

ccc a numeric vector

cohort a numeric vector

educ a numeric vector

educc a numeric vector

educr a numeric vector

inlfc a numeric vector

labforce a numeric vector

lfc a numeric vector

marst a numeric vector

marstc a numeric vector

marstr a numeric vector

nc a numeric vector

ncc a numeric vector

nchild a numeric vector

pc a numeric vector

pcc a numeric vector

wt a numeric vector

wtsupp a numeric vector

year a numeric vector

Details

test

Source

CPS

References

Luo and Hodges (2019)

Index