

6 ExpoCros module

6.1 Application domain

ExpoCros module was designed to analyze the cross-sectional data from exposome-wide association study (EWAS). This data structure can be obtained from the epidemiological designs of cross-section, case-control, and cohort. Users can easily get the modeling results and their visualization plots with high quality by following the detailed instructions in each step. For association, the model is chosen according to the data type of health outcome; while for prediction, most of the frequently-used models are evaluated for users' reference.

6.2 Theory

6.3 Work pipeline

Initialize package

Make sure that the required packages is already installed.

```
# The following two packages should be installed in advance
# devtools::install_github("ExposomeX/excros", force = TRUE)
# devtools::install_github("ExposomeX/extidy", force = TRUE)

# library(excros)
# library(extidy)
library(tidyverse)

# devtools::install_github("ExposomeX/exposomex", force = TRUE)
library(exposomex)
```

At first, you need to initialize the calculation environment using a series of initialization functions, e.g., InitCros, InitMo, InitTidy, InitViz, InitBiolink, etc. Here, we use the package “excros” for cross-sectional data analysis for example. The detailed information about the functions and returned value will be introduced in the following chapters.

```
res = InitCros()
res

## <eSet>
##   Public:
##     AddCommand: function (x)
##     AddLog: function (x)
##     clone: function (deep = FALSE)
##     EpiDesign: NULL
##     ExcecutionLog: Complete initializing the ExpoStat module.2022.12.14 14. ...
##     Expo: list
##     FileDirIn: NULL
##     FileDirOut: /home/ubuntu/@changxin/R_Exposome_1.0/output_144957LDYDKS
##     PID: 144957LDYDKS
##     RCommandLog: eSet <- InitCros(PID = Any ID your like, FileDirOut = An ...
##     VarsDel: NULL
```

Here, we can see that the returned value “res” is an R6 object. It contains an unique program ID of res\$PID (e.g., “100737GJMWJA”), which is random generated by the system. Users need use it in the following step for further data process.

Upload data

```
res1 <- LoadCros(res$PID,
                 UseExample = "example#1")

res1$Expo$Voca %>%
  dplyr::slice(1:20) %>%
  knitr::kable(format = "latex",
               align = "l") %>%
  kableExtra::kable_styling(full_width = F,
                             latex_options = "striped",
                             position = "left",
                             font_size = 10)
```

SerialNo	SerialNo_Raw	FullName	GroupName	Lod
Y1	Y1	Y_cont	Outcome	NA
Y2	Y2	Y_disc	Outcome	NA
C1	C1	Cov_1	Demography	NA
C2	C2	Cov_2	Demography	NA
C3	C3	Cov_3	Demography	NA
C4	C4	Cov_4	Demography	NA
C5	C5	Cov_5	Demography	NA
C6	C6	Cov_6	Demography	NA
X1	X1	TE_1	Chemical	0.5
X2	X2	TE_2	Chemical	0.5
X3	X3	TE_3	Chemical	0.5
X4	X4	TE_4	Chemical	0.5
X5	X5	TE_5	Chemical	0.5
X6	X6	TE_6	Chemical	0.5
X7	X7	TE_7	Chemical	0.5
X8	X8	TE_8	Chemical	0.5
X9	X9	CH1	Chemical	5.0
X10	X10	CH2	Chemical	5.0
X11	X11	CH3	Chemical	5.0
X12	X12	CH4	Chemical	5.0

```
res1$Expo$Data %>%
  dplyr::select(SampleID:C2) %>%
  dplyr::slice(1:20) %>%
  knitr::kable(format = "latex",
               align = "l") %>%
  kableExtra::kable_styling(full_width = F,
                             latex_options = "striped",
                             position = "left",
                             font_size = 10)
```

SampleID	SubjectID	Group	Y1	Y2	C1	C2
Tr1	S1	train	-101	1	26.86773	25.35056
Tr2	S2	train	-51	0	30.91822	23.94432
Tr3	S3	train	-37	0	25.82186	23.04579
Tr4	S4	train	-61	1	37.97640	21.21191
Tr5	S5	train	-28	0	31.64754	19.53762
Tr6	S6	train	-8	0	25.89766	20.77442
Tr7	S7	train	-63	1	32.43715	27.00009
Tr8	S8	train	-35	0	33.69162	22.13620
Tr9	S9	train	-14	0	32.87891	19.84672
Tr10	S10	train	-99	1	28.47306	29.60787
Tr11	S11	train	-60	0	37.55891	25.27530
Tr12	S12	train	-32	0	31.94922	23.28406
Tr13	S13	train	-73	0	26.89380	27.17545
Tr14	S14	train	-18	0	18.92650	26.65927
Tr15	S15	train	-48	0	35.62465	22.14227
Tr16	S16	train	-20	0	29.77533	30.61831
Tr17	S17	train	-9	0	29.91905	23.23492
Tr18	S18	train	-98	1	34.71918	19.72652
Tr19	S19	train	-70	0	34.10611	23.56680
Tr20	S20	train	-36	0	32.96951	24.62261

Tidy data

```
res2 <- DelMiss(res$PID)

res2$Expo$Voca %>%
  dplyr::slice(1:20) %>%
  knitr::kable(format = "latex",
               align = "l") %>%
  kableExtra::kable_styling(full_width = F,
                             latex_options = "striped",
                             position = "left",
                             font_size = 10)
```

SerialNo	SerialNo_Raw	FullName	GroupName	Lod
Y1	Y1	Y_cont	Outcome	NA
Y2	Y2	Y_disc	Outcome	NA
C1	C1	Cov_1	Demography	NA
C2	C2	Cov_2	Demography	NA
C3	C3	Cov_3	Demography	NA
C4	C4	Cov_4	Demography	NA
C5	C5	Cov_5	Demography	NA
C6	C6	Cov_6	Demography	NA
X1	X1	TE_1	Chemical	0.5
X2	X2	TE_2	Chemical	0.5
X3	X3	TE_3	Chemical	0.5
X4	X4	TE_4	Chemical	0.5
X5	X5	TE_5	Chemical	0.5
X6	X6	TE_6	Chemical	0.5
X7	X7	TE_7	Chemical	0.5
X8	X8	TE_8	Chemical	0.5
X9	X9	CH1	Chemical	5.0
X10	X10	CH2	Chemical	5.0
X11	X11	CH3	Chemical	5.0
X12	X12	CH4	Chemical	5.0

```
res2$Expo$Data %>%
  dplyr::select(SampleID:C2) %>%
  dplyr::slice(1:20) %>%
  knitr::kable(format = "latex",
               align = "l") %>%
  kableExtra::kable_styling(full_width = F,
                           latex_options = "striped",
                           position = "left",
                           font_size = 10)
```

SampleID	SubjectID	Group	Y1	Y2	C1	C2
Tr1	S1	train	-101	1	26.86773	25.35056
Tr2	S2	train	-51	0	30.91822	23.94432
Tr3	S3	train	-37	0	25.82186	23.04579
Tr4	S4	train	-61	1	37.97640	21.21191
Tr5	S5	train	-28	0	31.64754	19.53762
Tr6	S6	train	-8	0	25.89766	20.77442
Tr7	S7	train	-63	1	32.43715	27.00009
Tr8	S8	train	-35	0	33.69162	22.13620
Tr9	S9	train	-14	0	32.87891	19.84672
Tr10	S10	train	-99	1	28.47306	29.60787
Tr11	S11	train	-60	0	37.55891	25.27530
Tr12	S12	train	-32	0	31.94922	23.28406
Tr13	S13	train	-73	0	26.89380	27.17545
Tr14	S14	train	-18	0	18.92650	26.65927
Tr15	S15	train	-48	0	35.62465	22.14227
Tr16	S16	train	-20	0	29.77533	30.61831
Tr17	S17	train	-9	0	29.91905	23.23492
Tr18	S18	train	-98	1	34.71918	19.72652
Tr19	S19	train	-70	0	34.10611	23.56680
Tr20	S20	train	-36	0	32.96951	24.62261

Modeling

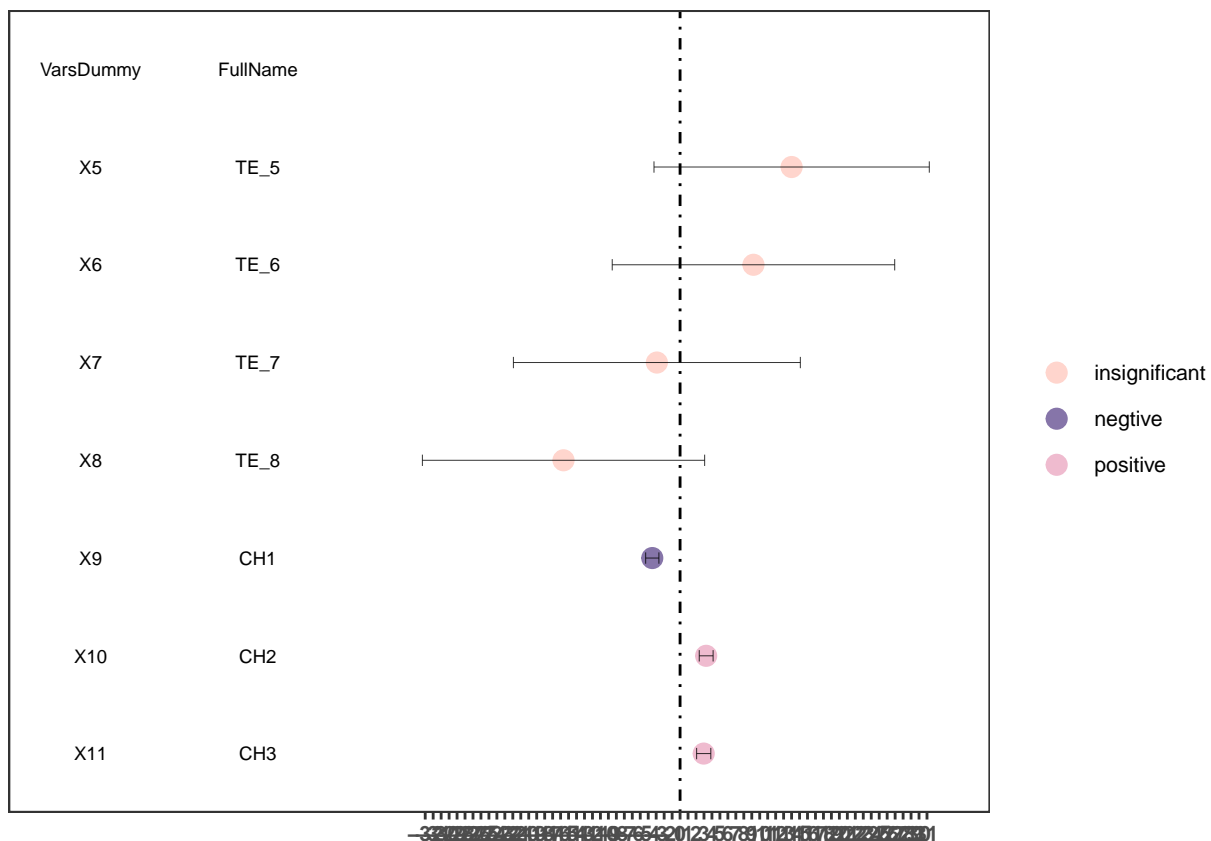
```
res3 = CrosAsso(PID = res$PID,
  EpiDesign = "cohort",
  VarsY = "Y1",
  VarsX = "X5,X6,X7,X8,X9,X10,X11",
  VarsN = "single.factor" ,
  VarsSel = F,
  VarsSelThr = 0.1,
  IncCova = T,
  Family = "gaussian",
  RepMsr = F,
  Corstr = "ar1")

res3$Y1_single.factor_cohort_gaussian %>%
  dplyr::select(SerialNo:ci_h) %>%
  dplyr::slice(1:20) %>%
  knitr::kable(format = "latex",
    align = "l") %>%
  kableExtra::kable_styling(full_width = F,
    latex_options = "striped",
    position = "left",
    font_size = 10)
```

SerialNo	Vars.dummy	beta.value	ci_l	ci_h
X5	X5	14.005401	-3.277425	31.288227
X6	X6	9.214098	-8.517261	26.945457
X7	X7	-2.915070	-20.927230	15.097090
X8	X8	-14.634061	-32.354215	3.086094
X9	X9	-3.489654	-4.319254	-2.660055
X10	X10	3.275350	2.404960	4.145739
X11	X11	2.968116	2.078754	3.857479

Visualize model

```
res4 = VizCrosAsso(PID = res$PID,
  VarsN="single.factor",
  VarsY="Y1",
  Layout = "forest",
  Brightness = "dark",
  Palette = "default1")
res4$Y1_single.factor_forest_dark_default1
```



```
FuncExit(PID = res$PID)

## [1] "Success to exit. Thanks for using ExposomeX platform!"
```