

# Variable Selection Result: Cox Proportional Hazards Model

Mingcheng Hu

## Table of contents

Table . . . . .	2
Minimally-adjusted Model . . . . .	4
Partially-adjusted Model . . . . .	6
Fully-adjusted Model . . . . .	8
Figure . . . . .	11

```
library(tidyverse)
library(kableExtra)

source("/work/users/y/u/yuukias/BIOS-Material/BIOS992/utils/csv_utils.r")
# * Don't use setwd() for Quarto documents!
# setwd("/work/users/y/u/yuukias/BIOS-Material/BIOS992/data")

variable_selection_results <- list()
adjust_type_str <- "minimal"
impute_type_str <- "unimputed"
include_statin <- "no"
file_path_minimal <- get_data_path("cox_var_select_result", adjust_type_str,
  ↪ impute_type_str, include_statin, model =
  "cox")
load(file_path_minimal)
variable_selection_results[[adjust_type_str]] <- variable_selection_matrix

adjust_type_str <- "partial"
impute_type_str <- "unimputed"
include_statin <- "no"
```

```

file_path_partial <- get_data_path("cox_var_select_result", adjust_type_str,
  ↪ impute_type_str, include_statin, model =
  "cox")
load(file_path_partial)
variable_selection_results[[adjust_type_str]] <- variable_selection_matrix

adjust_type_str <- "full"
impute_type_str <- "imputed"
include_statin <- "no"
file_path_full <- get_data_path("cox_var_select_result", adjust_type_str,
  ↪ impute_type_str, include_statin, model =
  "cox")
load(file_path_full)
variable_selection_results[[adjust_type_str]] <- variable_selection_matrix

```

## Table

```

visualize_table <- function(variable_selection_matrix) {
  variable_names_all <- rownames(variable_selection_matrix)
  symbol_selected <- "*"

  selection_table <- data.frame(
    Variable = variable_names_all,
    Univariate = ifelse(variable_selection_matrix[, "univariate"] == 1,
      ↪ symbol_selected, ""),
    Multivariate = ifelse(variable_selection_matrix[, "multivariate"] ==
      ↪ 1, symbol_selected, ""),
    LASSO = ifelse(variable_selection_matrix[, "lasso"] == 1,
      ↪ symbol_selected, ""),
    Stepwise = ifelse(variable_selection_matrix[, "stepwise"] == 1,
      ↪ symbol_selected, "")
  ) %>%
  mutate(Num_Selected = rowSums(variable_selection_matrix)) %>%
  arrange(desc(Num_Selected), Variable) %>%
  as.data.frame() %>%
  remove_rownames()

  variable_categories <- sapply(variable_names_all, determine_category)
  category_colors <- c(

```

```

    "covariate" = "#FFB6C1", #
    "time"      = "#1E90FF", #
    "frequency" = "#32CD32", #
    "poincare"  = "#FF4500", #
    "entropy"   = "#FF8C00", #
    "fractal"   = "#FFD700", #
    "unknown"   = "#000000" #
  )
  category_colors_names <- c(
    "covariate" = "pink", #
    "time"      = "blue", #
    "frequency" = "green", #
    "poincare"  = "red", #
    "entropy"   = "orange", #
    "fractal"   = "gold" #
  )
  category_legend <- sapply(names(category_colors_names), function(cat) {
    sprintf(
      "%s: %s",
      tools::toTitleCase(cat),
      tools::toTitleCase(category_colors_names[cat])
    )
  }) %>%
  paste(collapse = "; ")

selection_table %>%
  kbl(
    caption = "Variable Selection by Different Models",
    align = c("|l", "c", "c", "c", "c", "c|"),
    col.names = c("Variable", "Univariate", "Multivariate", "LASSO",
      ↪ "Stepwise", "Selected Times"),
    longtable = TRUE
  ) %>%
  kable_styling(
    bootstrap_options = c("striped", "hover", "condensed",
      ↪ "responsive"),
    position = "center",
    font_size = 9,
    latex_options = c("repeat_header", "striped", "HOLD_position")
  ) %>%
  # Add color for different categories of variables
  column_spec(1,

```

```

        color =
        ↪ category_colors[variable_categories[selection_table$Variable]],
        bold = TRUE
    ) %>%
    # Add a header colname for four columns: Univariate, Multivariate,
    ↪ LASSO, Stepwise
    add_header_above(c(
        " " = 1,
        "Selection Methods" = 4,
        " " = 1
    )) %>%
    footnote(
        general = sprintf("%s", category_legend),
        general_title = "Note:"
    )
}

```

## Minimally-adjusted Model

```

adjust_type_str <- "minimal"
impute_type_str <- "unimputed"
include_statin <- "no"
file_path_minimal <- get_data_path("cox_model_univariate", adjust_type_str,
    ↪ impute_type_str, include_statin, model =
    "cox")
load(file_path_minimal)

```

```
visualize_table(variable_selection_results[[adjust_type_str]])
```

Warning: 'xfun::attr()' is deprecated.  
 Use 'xfun::attr2()' instead.  
 See help("Deprecated")

Warning: 'xfun::attr()' is deprecated.  
 Use 'xfun::attr2()' instead.  
 See help("Deprecated")

Table 1: Variable Selection by Different Models

Variable	Selection Methods				Selected Times
	Univariate	Multivariate	LASSO	Stepwise	
HRV_CVI	*	*	*	*	4
HRV_CD	*	*		*	3
HRV_PI	*	*		*	3
HRV_PIP	*		*	*	3
HRV_RCMSEn	*	*		*	3
HRV_ApEn	*	*			2
HRV_CSI	*		*		2
HRV_LZC	*	*			2
HRV_SD1	*			*	2
HRV_AI	*				1
HRV_C1a	*				1
HRV_C1d	*				1
HRV_C2a	*				1
HRV_C2d	*				1
HRV_CMSEn	*				1
HRV_CSI_Modified	*				1
HRV_Ca	*				1
HRV_Cd	*				1
HRV_DFA_alpha1	*				1
HRV_FuzzyEn	*				1
HRV_GI	*				1
HRV_HFD	*				1
HRV_IALS	*				1
HRV_MFDFA_alpha1_Asymmetry	*				1
HRV_MFDFA_alpha1_Delta	*				1
HRV_MFDFA_alpha1_Max	*				1
HRV_MFDFA_alpha1_Peak	*				1
HRV_PAS	*				1
HRV_PSS	*				1
HRV_S		*			1
HRV_SD1SD2	*				1
HRV_SD1a	*				1
HRV_SD1d	*				1
HRV_SD2	*				1
HRV_SD2a	*				1
HRV_SD2d	*				1
HRV_SDNNa	*				1
HRV_SDNNd	*				1
HRV_SI	*				1
HRV_KFD					0
HRV_MFDFA_alpha1_Fluctuation					0
HRV_MFDFA_alpha1_Increment					0
HRV_MFDFA_alpha1_Mean					0
HRV_MFDFA_alpha1_Width					0
HRV_MSEn					0
HRV_ShanEn					0

Note:

Table 1: Variable Selection by Different Models (*continued*)

Variable	Univariate	Multivariate	LASSO	Stepwise	Selected Times
----------	------------	--------------	-------	----------	----------------

Covariate: Pink; Time: Blue; Frequency: Green; Poincare: Red; Entropy: Orange; Fractal: Gold

### Partially-adjusted Model

```
adjust_type_str <- "partial"
impute_type_str <- "unimputed"
include_statin <- "no"
file_path_partial <- get_data_path("cox_model_univariate", adjust_type_str,
  ↪ impute_type_str, include_statin, model =
  "cox")
load(file_path_partial)
```

```
visualize_table(variable_selection_results[[adjust_type_str]])
```

Warning: 'xfun::attr()' is deprecated.  
Use 'xfun::attr2()' instead.  
See help("Deprecated")

Warning: 'xfun::attr()' is deprecated.  
Use 'xfun::attr2()' instead.  
See help("Deprecated")

Table 2: Variable Selection by Different Models

Variable	Selection Methods				Selected Times
	Univariate	Multivariate	LASSO	Stepwise	
HRV_CD	*	*	*	*	4
HRV_HTI	*	*	*	*	4
HRV_PI	*	*	*	*	4
HRV_PIP	*	*	*	*	4
HRV_RCMSEn	*	*	*	*	4
HRV_ApEn	*	*	*		3
HRV_C2a	*		*	*	3
HRV_C2d	*		*	*	3
HRV_CSI	*		*	*	3
HRV_GI	*		*	*	3
HRV_IQRNN	*		*	*	3
HRV_MCVNN	*	*		*	3

Table 2: Variable Selection by Different Models (*continued*)

Variable	Univariate	Multivariate	LASSO	Stepwise	Selected Times
HRV_MadNN	*	*		*	3
HRV_MaxNN	*	*		*	3
HRV_MinNN	*		*	*	3
HRV_PAS	*		*	*	3
HRV_Prc20NN	*		*	*	3
HRV_SD1SD2	*		*	*	3
HRV_SDRMSSD	*		*	*	3
HRV_SI	*		*	*	3
HRV_AI	*			*	2
HRV_C1a	*			*	2
HRV_C1d	*			*	2
HRV_CSI_Modified	*			*	2
HRV_CVI	*			*	2
HRV_CVNN	*			*	2
HRV_CVSD	*			*	2
HRV_Cd	*			*	2
HRV_HFD	*			*	2
HRV_IALS	*			*	2
HRV_LF	*			*	2
HRV_LFn	*			*	2
HRV_LZC	*		*		2
HRV_MFDFA_alpha1_Max	*		*		2
HRV_MeanNN	*			*	2
HRV_MedianNN	*			*	2
HRV_PSS	*			*	2
HRV_Prc80NN	*			*	2
HRV_RMSSD	*			*	2
HRV_SD1	*			*	2
HRV_SD1a	*			*	2
HRV_SD1d	*			*	2
HRV_SD2	*			*	2
HRV_SD2a	*			*	2
HRV_SD2d	*			*	2
HRV_SDNN	*			*	2
HRV_SDS	*			*	2
HRV_TINN	*			*	2
HRV_VHF	*			*	2
HRV_pNN20	*			*	2
HRV_CMSEn	*				1
HRV_Ca	*				1
HRV_DFA_alpha1	*				1
HRV_FuzzyEn	*				1
HRV_HF				*	1
HRV_HFn				*	1
HRV_KFD			*		1
HRV_LFHF				*	1
HRV_LnHF				*	1
HRV_MFDFA_alpha1_Asymmetry	*				1

Table 2: Variable Selection by Different Models (*continued*)

Variable	Univariate	Multivariate	LASSO	Stepwise	Selected Times
HRV_MFDFA_alpha1_Delta	*				1
HRV_MFDFA_alpha1_Peak	*				1
HRV_S				*	1
HRV_SDNNa	*				1
HRV_SDNNd	*				1
HRV_TP				*	1
HRV_pNN50				*	1
HRV_MFDFA_alpha1_Fluctuation					0
HRV_MFDFA_alpha1_Increment					0
HRV_MFDFA_alpha1_Mean					0
HRV_MFDFA_alpha1_Width					0
HRV_MSEn					0
HRV_ShanEn					0

Note:

Covariate: Pink; Time: Blue; Frequency: Green; Poincare: Red; Entropy: Orange; Fractal: Gold

### Fully-adjusted Model

```
adjust_type_str <- "full"
impute_type_str <- "imputed"
include_statin <- "no"
file_path_full <- get_data_path("cox_model_univariate", adjust_type_str,
  ↪ impute_type_str, include_statin, model =
  "cox")
load(file_path_full)
```

```
visualize_table(variable_selection_results[[adjust_type_str]])
```

Warning: 'xfun::attr()' is deprecated.  
Use 'xfun::attr2()' instead.  
See help("Deprecated")

Warning: 'xfun::attr()' is deprecated.  
Use 'xfun::attr2()' instead.  
See help("Deprecated")



Table 3: Variable Selection by Different Models

Variable	Selection Methods				Selected Times
	Univariate	Multivariate	LASSO	Stepwise	
BMI	*	*	*	*	4
age	*	*	*	*	4
hdl_chol	*	*	*	*	4
HRV_HTI	*	*		*	3
max_workload	*	*		*	3
HRV_ApEn	*	*			2
HRV_FuzzyEn	*	*			2
HRV_Prc20NN	*			*	2
max_heart_rate	*		*		2
systolic_bp	*		*		2
HRV_AI	*				1
HRV_C1a	*				1
HRV_C1d	*				1
HRV_C2a	*				1
HRV_C2d	*				1
HRV_CD	*				1
HRV_CMSEn	*				1
HRV_CSI	*				1
HRV_CSI_Modified	*				1
HRV_CVI	*				1
HRV_CVNN	*				1
HRV_CVSD	*				1
HRV_Ca	*				1
HRV_Cd	*				1
HRV_DFA_alpha1	*				1
HRV_GI	*				1
HRV_HFD	*				1
HRV_IALS	*				1
HRV_IQRNN	*				1
HRV_LF	*				1
HRV_LFn	*				1
HRV_LZC	*				1
HRV_MCVNN	*				1
HRV_MFDFA_alpha1_Asymmetry	*				1
HRV_MFDFA_alpha1_Delta	*				1
HRV_MFDFA_alpha1_Max	*				1
HRV_MFDFA_alpha1_Peak	*				1
HRV_MadNN	*				1
HRV_MaxNN	*				1
HRV_MeanNN	*				1
HRV_MedianNN	*				1
HRV_MinNN	*				1
HRV_PAS	*				1
HRV_PI	*				1
HRV_PIP	*				1
HRV_PSS	*				1
HRV_Prc80NN	*				1

Table 3: Variable Selection by Different Models (*continued*)

Variable	Univariate	Multivariate	LASSO	Stepwise	Selected Times
HRV_RCMSEn	*				1
HRV_RMSSD	*				1
HRV_SD1	*				1
HRV_SD1SD2	*				1
HRV_SD1a	*				1
HRV_SD1d	*				1
HRV_SD2	*				1
HRV_SD2a	*				1
HRV_SD2d	*				1
HRV_SDNN	*				1
HRV_SDNNa	*				1
HRV_SDNNd	*				1
HRV_SDRMSSD	*				1
HRV_SDSd	*				1
HRV_SI	*				1
HRV_ShanEn		*			1
HRV_TINN	*				1
HRV_VHF	*				1
HRV_pNN20	*				1
total_chol	*				1
HRV_HF					0
HRV_HF <sub>n</sub>					0
HRV_KFD					0
HRV_LFHF					0
HRV_LnHF					0
HRV_MFDFA_alpha1_Fluctuation					0
HRV_MFDFA_alpha1_Increment					0
HRV_MFDFA_alpha1_Mean					0
HRV_MFDFA_alpha1_Width					0
HRV_MSEn					0
HRV_S					0
HRV_TP					0
HRV_pNN50					0
activity					0
diabetes					0
education					0
ethnicity					0
hypertension_treatment					0
sex					0
smoking					0

Note:

Covariate: Pink; Time: Blue; Frequency: Green; Poincare: Red; Entropy: Orange; Fractal: Gold

## Figure

```
# * Note the functions here to compose the dataframe
df_cox <- bind_rows(
  lapply(names(variable_selection_results), function(adjust_type_str) {
    as.data.frame(variable_selection_results[[adjust_type_str]]) %>%
      mutate(Feature =
        ↪ rownames(variable_selection_results[[adjust_type_str]])) %>%
      mutate(Adjustment = adjust_type_str) %>%
      pivot_longer(cols = -c(Feature, Adjustment), names_to = "Model",
        ↪ values_to = "Selected")
  })
)
df_cox # note now each row corresponds to a single model
```

```
# A tibble: 824 x 4
  Feature      Adjustment Model      Selected
  <chr>        <chr>      <chr>      <dbl>
1 HRV_SD1     minimal    univariate      1
2 HRV_SD1     minimal    multivariate    0
3 HRV_SD1     minimal    lasso           0
4 HRV_SD1     minimal    stepwise        1
5 HRV_SD2     minimal    univariate      1
6 HRV_SD2     minimal    multivariate    0
7 HRV_SD2     minimal    lasso           0
8 HRV_SD2     minimal    stepwise        0
9 HRV_SD1SD2  minimal    univariate      1
10 HRV_SD1SD2 minimal    multivariate    0
# i 814 more rows
```

```
# We would like to sum the selected times for each feature across different
  ↪ models
df_cox <- df_cox %>%
  group_by(Feature, Adjustment) %>%
  summarise(Selected = sum(Selected)) %>%
  ungroup()
```

`summarise()` has grouped output by 'Feature'. You can override using the  
`.groups` argument.

```
df_cox
```

```
# A tibble: 206 x 3
  Feature Adjustment Selected
  <chr>      <chr>      <dbl>
1 BMI      full        4
2 HRV_AI   full        1
3 HRV_AI   minimal      1
4 HRV_AI   partial      2
5 HRV_ApEn full        2
6 HRV_ApEn minimal    2
7 HRV_ApEn partial    3
8 HRV_C1a  full        1
9 HRV_C1a  minimal    1
10 HRV_C1a partial    2
# i 196 more rows
```

```
# However, there are too many features to display on a single plot, we will
  ↳ need to get category for each feature.
```

```
df_cox <- df_cox %>%
  mutate(Category = sapply(Feature, determine_category))
```

```
df_cox$Adjustment <- factor(df_cox$Adjustment, levels = c("minimal",
  ↳ "partial", "full"))
```

```
ggplot(df_cox, aes(x = Feature, y = Selected, fill = Adjustment)) +
  geom_bar(stat = "identity", position = "dodge") +
  scale_fill_brewer(palette = "Set2") +
  facet_wrap(~Category, scales = "free_x") +
  # * Rotate the labels
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

