

Correlations

Eduardo Yuki Yada

Imports

```
library(tidyverse)
library(yaml)
library(kableExtra)
library(ggcorrplot)
```

Loading data

```
load('../dataset/processed_data.RData')
load('../dataset/processed_dictionary.RData')

columns_list <- yaml.load_file("../auxiliar/columns_list.yaml")

outcome_column <- params$outcome_column
```

Functions

```
niceFormatting = function(df, caption="", digits = 2, font_size = NULL){
  df %>%
    kbl(booktabs = T, longtable = T, caption = caption, digits = digits, format = "latex") %>%
    kable_styling(font_size = font_size,
                  latex_options = c("striped", "HOLD_position", "repeat_header"))
}
```

Correlation

```
na_eligible_columns <- df %>%
  summarise(across(everything(), ~ mean(is.na(.)))) %>%
  select_if(function(.) last(.) < 0.8) %>%
  names

unique_eligible_columns <- df %>%
  summarise(across(everything(), ~ length(unique(.)))) %>%
  select_if(function(.) last(.) > 1) %>%
  names

pre_columns = df_names %>%
  filter(momento.aquisicao == 'Admissão t0') %>%
  .$variable.name

weird_columns <- c('dieta_parentral', 'dieta_enteral')

eligible_columns <- intersect(na_eligible_columns,
                             unique_eligible_columns) %>%
  intersect(pre_columns)
```

```

eligible_columns <- setdiff(eligible_columns, weird_columns)

corr <- df %>%
  select(all_of(intersect(columns_list$numerical_columns,
                          eligible_columns))) %>%

  drop_na %>%
  cor %>%
  as.matrix

## Warning in cor(.): the standard deviation is zero

corr_table <- corr %>%
  as.data.frame %>%
  tibble::rownames_to_column(var = 'row') %>%
  tidyr::pivot_longer(-row, names_to = 'column', values_to = 'correlation') %>%
  filter(row < column)

rename_column <- function(df, column_name){
  variable.name <- 'variable.name'
  df <- df %>%
    left_join(df_names %>% select(variable.name, abbrev.field.label),
              by = setNames(variable.name, column_name)) %>%
    select(-all_of(column_name)) %>%
    rename(!sym(column_name) := abbrev.field.label) %>%
    relocate(!sym(column_name))
}

corr_table %>%
  filter(correlation > 0.9) %>%
  rename_column('row') %>%
  rename_column('column') %>%
  select(row, column, correlation) %>%
  niceFormatting(caption = "Pearson Correlation", font_size = 9)

```

Table 1: Pearson Correlation

row	column	correlation
Idade no momento do primeiro procedimento	Idade no Procedimento 1	1.00
Núm. de hospitalizações pré-procedimento	Número da Admissão T0	0.98
Ano da admissão T0	Ano do procedimento 1	1.00
Antibióticos	Quantidade de antimicrobianos	1.00
Quantidade de classes medicamentosas de ação cardiovascular	Quantidade de classes medicamentosas utilizadas	0.91
Quantidade de procedimentos invasivos	Suporte cardiocirculatório	0.97
ECG	Quantidade de exames por métodos gráficos	1.00
Exames laboratoriais	Radiografias	0.90
Quantidade de exames de análises clínicas	Exames laboratoriais	1.00
Quantidade de exames de análises clínicas	Radiografias	0.90
Quantidade de exames de análises clínicas	Quantidade de exames diagnóstico por imagem	0.93
Quantidade de exames diagnóstico por imagem	Exames laboratoriais	0.93
Quantidade de exames diagnóstico por imagem	Radiografias	0.98

Hypothesis Tests

```

df_wilcox <- tibble()

for (variable in columns_list$numerical_columns){
  if (mean(is.na(df[[variable]])) > 0.95) next

  x <- filter(df, !!sym(outcome_column) == 0)[[variable]]
  y <- filter(df, !!sym(outcome_column) == 1)[[variable]]

```

```

test = tryCatch(wilcox.test(x, y, alternative = "two.sided", exact = FALSE),
  error=function(cond) {
    message("Can't calculate Wilcox test for variable ", variable)
    message(cond)
    return(list(statistic = NaN, p.value = NaN))
  })

df_wilcox = bind_rows(df_wilcox,
  list("Variable" = variable,
    "Statistic" = test$statistic,
    "p-value" = test$p.value))
}

df_wilcox <- df_wilcox %>%
  arrange(`p-value`) %>%
  mutate(`Statistic` = round(`Statistic`, 3)) %>%
  rename_column('Variable')

significant_numerical_columns <- df_wilcox %>%
  filter(`p-value` <= 0.25) %>%
  select(Variable) %>%
  pull

df_wilcox %>%
  mutate(`p-value` = case_when(`p-value` == 1 ~ sprintf('> 0%s999', getOption("OutDec")),
    `p-value` < 0.001 ~ sprintf('< 0%s001', getOption("OutDec")),
    TRUE ~ as.character(round(`p-value`, 3)))) %>%
  niceFormatting(caption = "Mann-Whitney Test")

```

Table 2: Mann-Whitney Test

Variable	Statistic	p-value
Número de atendimentos	3860198.0	< 0.001
Núm. de hospitalizações pós-procedimento	3238842.0	< 0.001
Readmissão em até 30 dias	6439005.0	< 0.001
Readmissão entre 31 a 60 dias	4236952.5	< 0.001
Readmissão entre 61 a 180 dias	0.0	< 0.001
Readmissão em até 1 ano	389373.5	< 0.001
Número de procedimentos em até 180 dias	9616368.5	< 0.001
Número de procedimentos em até 60 dias	9893287.5	< 0.001
Número de procedimentos em até 30 dias	10003042.0	< 0.001
Óbito durante algum episódio de readmissão hospitalar	8338963.5	< 0.001
Tempo entre o P1 e P2	2164258.0	< 0.001
Óbito	8778238.5	< 0.001
Quantidade de classes medicamentosas utilizadas	3505150.5	< 0.001
Número da Admissão T0	8743766.0	< 0.001
Quantidade de medicamentos de ação cardiovascular	4886965.5	< 0.001
Quantidade de exames diagnóstico por imagem	5656440.0	< 0.001
Antiarrítmicos	5551729.0	< 0.001
Quantidade de classes medicamentosas de ação cardiovascular	3007519.5	< 0.001
DVA	5489048.0	< 0.001
Quantidade de exames por métodos gráficos	5825362.5	< 0.001
Radiografias	5860552.5	< 0.001
UTI durante a admissão T0	9262145.0	< 0.001
ECG	5873146.5	< 0.001
Antagonista da Aldosterona	5609444.5	< 0.001
Equipe Multiprofissional	6032726.5	< 0.001

Table 2: Mann-Whitney Test (*continued*)

Variable	Statistic	p-value
Insuficiência cardíaca	5554513.0	< 0.001
Exames laboratoriais	5910126.5	< 0.001
Quantidade de exames de análises clínicas	5910722.5	< 0.001
Diuretico	5370033.0	< 0.001
Ultrassom	6780198.5	< 0.001
Núm. de hospitalizações pré-procedimento	9197237.0	< 0.001
Transplante cardíaco	7431750.0	< 0.001
Ecocardiograma	6340792.0	< 0.001
Biopsias	7366317.5	< 0.001
Tempo de seguimento total	12611007.0	< 0.001
Quantidade de procedimentos invasivos	6659612.0	< 0.001
Culturas	6770465.0	< 0.001
Número de comorbidades	9225813.0	< 0.001
Ano do procedimento 2	1583112.0	< 0.001
Quantidade de exames histopatológicos	7315755.0	< 0.001
Cateterismo	6898637.5	< 0.001
Número de procedimentos em até 1 ano	10561649.0	< 0.001
Anticoagulantes orais	6278301.5	< 0.001
Ressonancia magnetica	7051355.5	< 0.001
Holter	6951055.5	< 0.001
Tempo de sobrevida	465881.5	< 0.001
Psicofármacos	5765074.0	< 0.001
Vasodilator	5860700.5	< 0.001
Antiviral	6588344.0	< 0.001
Tomografia	6973424.5	< 0.001
Cateter venoso central	7248482.5	< 0.001
Cintilografia	7182434.0	< 0.001
Quantidade de antimicrobianos	5776176.0	< 0.001
Antibióticos	5782587.5	< 0.001
Estatinas	5928379.0	< 0.001
Digoxina	6364111.0	< 0.001
Bloqueador do canal de calcio	6505131.5	< 0.001
Exames endoscópicos	7353950.0	< 0.001
Antifúngicos	6485399.0	< 0.001
Ano do procedimento 3	193655.5	< 0.001
Betabloqueador	6272007.0	< 0.001
IECA/BRA	5949670.0	< 0.001
Eletrofisiologia	7276612.0	< 0.001
Diárias no serviço de Emergência na admissão T0	4061491.5	< 0.001
Tempo entre o P2 e P3	185218.5	< 0.001
Idade no Procedimento 2	1447543.0	< 0.001
Antiplaquetario EV	6618989.5	< 0.001
Óbito hospitalar	11016366.5	< 0.001
Outros procedimentos cirúrgicos	7268544.5	< 0.001
Bomba de infusão contínua	6460618.0	< 0.001
Instalação de CEC	7416797.0	< 0.001
Citologias	7452001.5	< 0.001
Suporte cardiocirculatório	7482157.5	< 0.001
Insulina	6476057.0	< 0.001
Idade no momento do primeiro procedimento	11551514.0	< 0.001
Idade no Procedimento 1	11551514.0	< 0.001
Intervenção coronária percutânea	7450866.0	< 0.001
Transfusão de hemoderivados	7448072.5	< 0.001

Table 2: Mann-Whitney Test (*continued*)

Variable	Statistic	p-value
Anticonvulsivante	6566585.0	< 0.001
Diálise durante a admissão T0	10771004.0	< 0.001
Angio RM	7498942.5	< 0.001
Espirometria / Ergoespirometria	7493516.0	0.003
Cardioversão/ Desfibrilação	6561262.5	0.004
Angio TC	7434508.0	0.005
Interconsulta médica	7337473.0	0.006
Tilt Test	7510324.5	0.013
Antiretroviral	6688603.5	0.014
Teste de esforço	7485620.0	0.019
Flebografia	7469549.5	0.02
PET-CT	7500402.5	0.021
Intervenção cardiovascular em laboratório de hemodinâmica	7503339.0	0.026
Ventilação não invasiva	7561748.0	0.043
Número de Mudanças do tipo de DCEI	1239082.0	0.054
Angioplastia	7522048.0	0.065
Marca-passo temporário	6563290.5	0.1
Arteriografia	7529019.0	0.133
Número de procedimentos na admissão T0	10774821.0	0.183
Ano da admissão T0	10984270.5	0.229
Dieta enteral	6582918.5	0.232
Ano do procedimento 1	11016782.0	0.25
Cirurgia Toracica	7524811.0	0.264
Óbito intraoperatório	10829497.0	0.288
Drenagem de tórax e punção pericárdica ou pleural	7552528.0	0.292
Polissonografia	7528668.5	0.297
Aortografia	7527061.0	0.303
Trombolítico	6696043.0	0.324
Idade no Procedimento 3	157236.0	0.373
Dieta parenteral	6599298.0	0.393
Cirurgia Cardiovascular	7567472.5	0.468
Antihipertensivo	6675546.0	0.516
Hipoglicemiante	6678665.5	0.576
Cavografia	7527137.0	0.601
Traqueostomia	7540049.0	0.71
Angiografia	7533698.5	0.719
Stent	7537884.0	0.747
Antiplaquetario VO	6702032.0	NaN
Hormonio tireoidiano	6702032.0	NaN
Broncodilator	6702032.0	NaN

```
df_chisq <- tibble()

for (variable in columns_list$categorical_columns){
  if (length(unique(df[[variable]])) > 1){
    test <- tryCatch(chisq.test(df[[outcome_column]],
                              df[[variable]] %>% replace_na('NA'), # counting NA as cat
                              simulate.p.value = TRUE),
                    error = function (cond) {
                      message("Can't calculate Chi Squared test for variable ", variable)
                      message(cond)
                      return(list(statistic = NaN, p.value = NaN))
                    })
  }
}
```

```

df_chisq <- bind_rows(df_chisq,
  list("Variable" = variable,
       "Statistic" = test$statistic,
       "p-value" = test$p.value))
}
}

df_chisq %>%
  arrange(`p-value`) %>%
  mutate(`p-value` = case_when(`p-value` == 1 ~ sprintf('> 0%s999', getOption("OutDec")),
                                `p-value` < 0.001 ~ sprintf('< 0%s001', getOption("OutDec")),
                                TRUE ~ as.character(round(`p-value`, 3))),
        `Statistic` = round(`Statistic`, 3)) %>%
  rename_column('Variable') %>%
  niceFormatting(caption = "Chi-squared test")

```

Table 3: Chi-squared test

Variable	Statistic	p-value
Escolaridade	40.01	< 0.001
Doença cardíaca	64.20	< 0.001
Doença cardíaca	34.21	< 0.001
Classe funcional de IC	47.38	< 0.001
Infarto do miocárdio prévio / Doença arterial coronariana	28.39	< 0.001
Insuficiência cardíaca	136.36	< 0.001
Fibrilação / flutter atrial	15.94	< 0.001
Parada cardíaca prévia/ Taquicardia ventricular instável	19.41	< 0.001
Transplante cardíaco prévio	20.97	< 0.001
Valvopatias/ Prótese valvares	15.33	< 0.001
Número de procedimentos	157.84	< 0.001
Tipo de Procedimento 1	121.85	< 0.001
Tipo de Reoperação 1	139.58	< 0.001
Tipo de Dispositivo ao final do procedimento 1	218.03	< 0.001
Tipo de Reoperação 2	889.43	< 0.001
Tipo de Dispositivo ao final do procedimento 2	119.26	< 0.001
Óbito intraoperatório 2	68.94	< 0.001
Tipo de Reoperação 3	93.43	< 0.001
Tipo de Dispositivo ao final do procedimento 3	108.42	< 0.001
Óbito intraoperatório 3	105.24	< 0.001
Tipo de Reoperação 4	105.03	< 0.001
Tipo de Dispositivo ao final do procedimento 4	104.29	< 0.001
Óbito intraoperatório 4	96.75	< 0.001
Óbito intraoperatório 5	14.37	< 0.001
Mudança do tipo de DCEI: entre o Procedimento 1 e Procedimento 2	67.45	< 0.001
Mudança do tipo de DCEI: entre o Procedimento 2 e Procedimento 3	104.90	< 0.001
Mudança do tipo de DCEI: entre o Procedimento 3 e Procedimento 4	97.38	< 0.001
UTI durante os episódios de hospitalização	215.54	< 0.001
Admissão em até 180 dias antes da T0	168.52	< 0.001
Readmissões pós-T0 com diárias de UTI	345.93	< 0.001
Desfecho principal da admissão T0	26.86	< 0.001
Readmissão cirúrgica em até 30 dias	1348.87	< 0.001
Readmissão cirúrgica entre 31 a 60 dias	913.13	< 0.001
Readmissão cirúrgica entre 61 a 180 dias	1468.13	< 0.001
Readmissão cirúrgica em até 1 ano	35.29	< 0.001
Desfecho final do estudo	342.62	< 0.001
Causa do óbito	351.08	< 0.001
Ventilação mecânica / IOT	72.38	< 0.001

Table 3: Chi-squared test (*continued*)

Variable	Statistic	p-value
Tipo de Reoperação 5	19.22	< 0.001
Sexo	10.30	0.001
Diabetes mellitus	11.26	0.001
Tipo de Dispositivo ao final do procedimento 5	19.81	0.001
Óbito intraoperatório 6	10.57	0.001
Mudança do tipo de DCEI: entre o Procedimento 4 e Procedimento 5	17.97	0.002
Diálise durante os episódios de hospitalização	10.49	0.002
Mudança do tipo de DCEI: entre o Procedimento 5 e Procedimento 6	14.12	0.003
Tipo de Reoperação 6	13.77	0.009
Tipo de Dispositivo ao final do procedimento 6	15.50	0.013
Hemodiálise	8.45	0.014
Estado de residência	52.50	0.014
Tipo de Dispositivo ao final do procedimento 7	15.13	0.017
Tipo de Reoperação 8	12.64	0.023
Insuficiência renal crônica	5.24	0.023
Tipo de Reoperação 7	11.45	0.023
Óbito intraoperatório 7	6.02	0.023
Endocardite prévia	4.47	0.033
Tipo de Dispositivo ao final do procedimento 8	11.65	0.034
Mudança do tipo de DCEI: entre o Procedimento 6 e Procedimento 7	6.96	0.039
Readmissões pós-T0 com diálise	10.92	0.062
Doença pulmonar obstrutiva crônica	3.21	0.073
Neoplasia em tratamento ou tratada recentemente	1.93	0.176
Acidente Vascular Cerebral/ Acidente isquêmico transitório prévios	0.87	0.395
Hipertensão arterial	0.33	0.554
Raça	4.38	0.617
Óbito intraoperatório 8	0.78	0.623
Óbito intraoperatório 1	0.72	0.661
Mudança do tipo de DCEI: entre o Procedimento 7 e Procedimento 8	1.13	0.661
Tipo de Reoperação 9	0.51	> 0.999
Tipo de Dispositivo ao final do procedimento 9	0.51	> 0.999
Óbito intraoperatório 9	0.51	> 0.999
Tipo de Reoperação 10	0.10	> 0.999
Tipo de Dispositivo ao final do procedimento 10	0.10	> 0.999
Óbito intraoperatório 10	0.10	> 0.999
Mudança do tipo de DCEI: entre o Procedimento 8 e Procedimento 9	0.51	> 0.999
Mudança do tipo de DCEI: entre o Procedimento 9 e Procedimento 10	0.10	> 0.999