

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	4657034.5	< 0.001
Núm. de hospitalizações pós-procedimento	3994063.5	< 0.001
Readmissão em até 30 dias	9871454.0	< 0.001
Readmissão entre 31 a 60 dias	7748636.0	< 0.001
Readmissão entre 61 a 180 dias	3664138.0	< 0.001
Readmissão em até 1 ano	0.0	< 0.001
Óbito durante algum episódio de readmissão hospitalar	10760936.0	< 0.001
Número de procedimentos em até 180 dias	12927064.5	< 0.001
Número de procedimentos em até 1 ano	13097200.5	< 0.001
Tempo entre o P1 e P2	2740769.5	< 0.001
Número de procedimentos em até 60 dias	13202308.5	< 0.001
Número de procedimentos em até 30 dias	13308402.0	< 0.001
Óbito	11334508.0	< 0.001
Número da Admissão T0	11160003.0	< 0.001
Quantidade de classes medicamentosas utilizadas	4581392.5	< 0.001
Quantidade de medicamentos de ação cardiovascular	6411969.5	< 0.001
Quantidade de exames diagnóstico por imagem	7426411.5	< 0.001
Quantidade de classes medicamentosas de ação cardiovascular	3896761.5	< 0.001
Núm. de hospitalizações pré-procedimento	11672486.0	< 0.001
Antiarrítmicos	7296806.5	< 0.001
Quantidade de exames por métodos gráficos	7542501.5	< 0.001
DVA	7167455.5	< 0.001
ECG	7618335.5	< 0.001
Equipe Multiprofissional	7818573.0	< 0.001
Insuficiência cardíaca	7188500.5	< 0.001

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

Variable	Statistic	p-value
Antagonista da Aldosterona	7273831.5	< 0.001
UTI durante a admissão T0	12176855.0	< 0.001
Diuretico	6937008.0	< 0.001
Exames laboratoriais	7701961.5	< 0.001
Quantidade de exames de análises clínicas	7702891.0	< 0.001
Radiografias	7751583.0	< 0.001
Ultrassom	8841632.0	< 0.001
Ecocardiograma	8253132.0	< 0.001
Holter	8870337.5	< 0.001
Quantidade de procedimentos invasivos	8684567.0	< 0.001
Biopsias	9550336.5	< 0.001
Tempo de seguimento total	16124772.5	< 0.001
Número de comorbidades	12156587.5	< 0.001
Transplante cardíaco	9632924.5	< 0.001
Ano do procedimento 2	2003170.0	< 0.001
Ressonancia magnetica	9112217.5	< 0.001
Cateterismo	8964854.0	< 0.001
Psicofármacos	7484463.5	< 0.001
Anticoagulantes orais	8161569.0	< 0.001
Culturas	8904264.0	< 0.001
Quantidade de exames histopatológicos	9503325.5	< 0.001
Tomografia	9061021.0	< 0.001
Vasodilator	7691487.0	< 0.001
Cateter venoso central	9394908.5	< 0.001
Cintilografia	9319038.0	< 0.001
Estatinas	7660020.5	< 0.001
Ano do procedimento 3	245761.5	< 0.001
Quantidade de antimicrobianos	7545156.5	< 0.001
Antibióticos	7562438.0	< 0.001
Digoxina	8240059.0	< 0.001
Bloqueador do canal de calcio	8418223.0	< 0.001
Antiviral	8542831.5	< 0.001
Tempo de sobrevida	536608.5	< 0.001
Eletrofisiologia	9369088.0	< 0.001
IECA/BRA	7725944.0	< 0.001
Exames endoscópicos	9534421.0	< 0.001
Bomba de infusão contínua	8316601.5	< 0.001
Tempo entre o P2 e P3	233898.0	< 0.001
Antifúngicos	8420932.0	< 0.001
Idade no Procedimento 2	1864696.5	< 0.001
Óbito hospitalar	14359644.0	< 0.001
Diárias no serviço de Emergência na admissão T0	5219539.5	< 0.001
Betabloqueador	8227091.0	< 0.001
Instalação de CEC	9589283.0	< 0.001
Outros procedimentos cirúrgicos	9427237.0	< 0.001
Intervenção coronária percutânea	9625714.5	< 0.001
Suporte cardiocirculatório	9673704.5	< 0.001
Antiplaquetario EV	8577459.5	< 0.001
Idade no momento do primeiro procedimento	14988108.0	< 0.001
Idade no Procedimento 1	14988108.0	< 0.001
Espirometria / Ergoespirometria	9669024.0	< 0.001
Transfusão de hemoderivados	9628706.5	< 0.001
Angio RM	9689411.5	< 0.001

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

Variable	Statistic	p-value
Citologias	9660226.0	< 0.001
Número de Mudanças do tipo de DCEI	1557590.0	< 0.001
Angio TC	9582022.5	< 0.001
Cardioversão/ Desfibrilação	8485049.5	< 0.001
Insulina	8452933.5	< 0.001
Angioplastia	9709113.5	< 0.001
Arteriografia	9720576.5	0.002
Intervenção cardiovascular em laboratório de hemodinâmica	9686577.5	0.002
Anticonvulsivante	8529225.0	0.003
Diálise durante a admissão T0	14048228.5	0.004
Dieta parenteral	8521111.0	0.016
Ano da admissão T0	14470565.0	0.028
Tilt Test	9713421.0	0.028
Interconsulta médica	9561432.0	0.029
Flebografia	9668817.5	0.031
Ano do procedimento 1	14509474.0	0.033
Teste de esforço	9689764.0	0.043
PET-CT	9704773.0	0.05
Ventilação não invasiva	9766558.0	0.057
Antiretroviral	8647147.5	0.063
Marca-passo temporário	8494219.0	0.07
Aortografia	9722463.5	0.111
Dieta enteral	8515948.5	0.117
Cirurgia Toracica	9721714.0	0.141
Óbito intraoperatório	14107138.0	0.209
Polissonografia	9728990.5	0.224
Idade no Procedimento 3	195050.0	0.279
Número de procedimentos na admissão T0	14053235.5	0.282
Traqueostomia	9733096.0	0.395
Antihipertensivo	8623535.0	0.447
Cirurgia Cardiovascular	9772230.0	0.5
Trombolítico	8654989.0	0.584
Hipoglicemiante	8632775.0	0.584
Stent	9741204.0	0.703
Drenagem de tórax e punção pericárdica ou pleural	9734407.5	0.718
Angiografia	9742356.0	0.859
Cavografia	9738335.5	0.927
Antiplaquetario VO	8658767.0	NaN
Hormonio tireoidiano	8658767.0	NaN
Broncodilator	8658767.0	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
Sexo	23.15	< 0.001
Escolaridade	51.50	< 0.001
Doença cardíaca	97.97	< 0.001
Doença cardíaca	46.32	< 0.001
Classe funcional de IC	64.73	< 0.001
Infarto do miocárdio prévio / Doença arterial coronariana	30.86	< 0.001
Insuficiência cardíaca	189.25	< 0.001
Fibrilação / flutter atrial	14.24	< 0.001
Parada cardíaca prévia/ Taquicardia ventricular instável	18.71	< 0.001
Valvopatias/ Prótese valvares	19.95	< 0.001
Número de procedimentos	203.58	< 0.001
Tipo de Procedimento 1	129.22	< 0.001
Tipo de Reoperação 1	153.74	< 0.001
Tipo de Dispositivo ao final do procedimento 1	290.68	< 0.001
Tipo de Reoperação 2	1076.23	< 0.001
Tipo de Dispositivo ao final do procedimento 2	206.26	< 0.001
Óbito intraoperatório 2	91.97	< 0.001
Tipo de Reoperação 3	115.22	< 0.001
Tipo de Dispositivo ao final do procedimento 3	145.02	< 0.001
Óbito intraoperatório 3	133.08	< 0.001
Tipo de Reoperação 4	148.82	< 0.001
Tipo de Dispositivo ao final do procedimento 4	143.20	< 0.001
Óbito intraoperatório 4	140.25	< 0.001
Tipo de Reoperação 5	40.78	< 0.001
Tipo de Dispositivo ao final do procedimento 5	38.15	< 0.001
Óbito intraoperatório 5	37.35	< 0.001
Tipo de Dispositivo ao final do procedimento 7	33.27	< 0.001
Mudança do tipo de DCEI: entre o Procedimento 1 e Procedimento 2	96.08	< 0.001
Mudança do tipo de DCEI: entre o Procedimento 2 e Procedimento 3	134.00	< 0.001
Mudança do tipo de DCEI: entre o Procedimento 3 e Procedimento 4	140.36	< 0.001
Mudança do tipo de DCEI: entre o Procedimento 4 e Procedimento 5	38.20	< 0.001
Mudança do tipo de DCEI: entre o Procedimento 5 e Procedimento 6	20.96	< 0.001
Diálise durante os episódios de hospitalização	19.28	< 0.001
UTI durante os episódios de hospitalização	294.83	< 0.001
Admissão em até 180 dias antes da T0	244.84	< 0.001
Readmissões pós-T0 com diárias de UTI	435.46	< 0.001
Readmissões pós-T0 com diálise	27.18	< 0.001
Desfecho principal da admissão T0	37.65	< 0.001

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

Variable	Statistic	p-value
Readmissão cirúrgica em até 30 dias	962.32	< 0.001
Readmissão cirúrgica entre 31 a 60 dias	651.45	< 0.001
Readmissão cirúrgica entre 61 a 180 dias	1047.41	< 0.001
Readmissão cirúrgica em até 1 ano	926.91	< 0.001
Desfecho final do estudo	478.75	< 0.001
Causa do óbito	490.85	< 0.001
Ventilação mecânica / IOT	83.63	< 0.001
Transplante cardíaco prévio	20.20	< 0.001
Insuficiência renal crônica	10.95	< 0.001
Mudança do tipo de DCEI: entre o Procedimento 6 e Procedimento 7	17.81	< 0.001
Óbito intraoperatório 6	15.81	0.001
Óbito intraoperatório 7	13.90	0.001
Estado de residência	61.25	0.002
Diabetes mellitus	8.44	0.003
Tipo de Reoperação 6	14.99	0.006
Tipo de Dispositivo ao final do procedimento 6	17.38	0.006
Tipo de Reoperação 7	17.01	0.006
Hemodiálise	7.44	0.015
Raça	14.72	0.036
Endocardite prévia	3.92	0.042
Tipo de Reoperação 8	8.02	0.06
Doença pulmonar obstrutiva crônica	3.05	0.075
Tipo de Dispositivo ao final do procedimento 8	7.55	0.124
Mudança do tipo de DCEI: entre o Procedimento 7 e Procedimento 8	2.32	0.251
Acidente Vascular Cerebral/ Acidente isquêmico transitório prévios	1.33	0.289
Neoplasia em tratamento ou tratada recentemente	1.01	0.335
Óbito intraoperatório 8	1.69	0.382
Tipo de Dispositivo ao final do procedimento 9	1.47	0.512
Óbito intraoperatório 1	1.01	0.615
Hipertensão arterial	0.23	0.656
Tipo de Reoperação 9	0.25	> 0.999
Óbito intraoperatório 9	0.25	> 0.999
Tipo de Reoperação 10	0.14	> 0.999
Tipo de Dispositivo ao final do procedimento 10	0.14	> 0.999
Óbito intraoperatório 10	0.14	> 0.999
Mudança do tipo de DCEI: entre o Procedimento 8 e Procedimento 9	0.25	> 0.999
Mudança do tipo de DCEI: entre o Procedimento 9 e Procedimento 10	0.14	> 0.999