

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
Óbito intraoperatório	0.0	< 0.001
Óbito hospitalar	1380.5	< 0.001
Aortografia	46476.5	< 0.001
Óbito	15328.5	< 0.001
Dieta enteral	43569.0	< 0.001
Cardioversão/ Desfibrilação	43606.5	< 0.001
Óbito durante algum episódio de readmissão hospitalar	63195.5	< 0.001
DVA	25327.0	0.001
Cirurgia Cardiovascular	41800.0	0.001
Número de Mudanças do tipo de DCEI	5693.0	0.003
Instalação de CEC	47040.5	0.004
Tempo de seguimento total	124324.5	0.018
Bomba de infusão contínua	44277.0	0.026
Ano do procedimento 2	14795.5	0.073
Tempo de sobrevida	14475.5	0.077
IECA/BRA	66579.5	0.09
Idade no Procedimento 2	14381.0	0.101
Holter	44881.0	0.139
Quantidade de procedimentos invasivos	42678.5	0.149
Antagonista da Aldosterona	60376.0	0.151
Quantidade de medicamentos de ação cardiovascular	36194.5	0.173
Idade no momento do primeiro procedimento	108452.5	0.184
Idade no Procedimento 1	108452.5	0.184
Cateterismo	45444.0	0.185
Readmissão em até 1 ano	99137.5	0.209

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

Variable	Statistic	p-value
Culturas	60432.0	0.255
Radiografias	41315.5	0.269
Estatinas	59550.0	0.274
Número de comorbidades	72058.0	0.276
Readmissão entre 61 a 180 dias	96261.0	0.288
Ecocardiograma	43421.0	0.292
Interconsulta médica	46704.0	0.299
Idade no Procedimento 3	3761.0	0.344
Tomografia	58264.0	0.348
Número de atendimentos	74736.5	0.358
Quantidade de classes medicamentosas de ação cardiovascular	34247.0	0.362
Quantidade de exames diagnóstico por imagem	43611.5	0.379
Equipe Multiprofissional	44594.5	0.387
Ano do procedimento 1	101211.0	0.391
Ano da admissão T0	101020.0	0.393
Readmissão entre 31 a 60 dias	93054.5	0.417
Insulina	53384.0	0.445
Diárias no serviço de Emergência na admissão T0	27582.5	0.452
Outros procedimentos cirúrgicos	56404.0	0.459
Ressonancia magnetica	49729.5	0.47
Núm. de hospitalizações pré-procedimento	79388.5	0.481
UTI durante a admissão T0	80558.5	0.484
Digoxina	52660.0	0.502
Anticoagulantes orais	52616.0	0.506
Readmissão em até 30 dias	91388.0	0.512
Diuretico	55762.5	0.533
Hipoglicemiante	52168.0	0.546
Antihipertensivo	52060.0	0.557
Exames laboratoriais	46861.0	0.565
Quantidade de exames de análises clínicas	46899.5	0.567
Eletrofisiologia	55020.0	0.574
Cintilografia	54992.0	0.576
Anticonvulsivante	51852.0	0.578
Quantidade de exames por métodos gráficos	47308.0	0.594
Núm. de hospitalizações pós-procedimento	94908.5	0.614
Psicofármacos	45561.5	0.63
Tempo entre o P2 e P3	3360.0	0.633
Angio TC	54456.0	0.635
Cateter venoso central	54296.0	0.654
Antifúngicos	51180.0	0.659
Ultrassom	50695.5	0.664
Ano do procedimento 3	3316.5	0.669
ECG	48503.0	0.674
Número de procedimentos na admissão T0	89325.5	0.69
Bloqueador do canal de calcio	50868.0	0.705
Flebografia	53920.0	0.707
Número de procedimentos em até 180 dias	88995.5	0.733
Quantidade de classes medicamentosas utilizadas	35613.0	0.736
Número da Admissão T0	84266.5	0.74
Antiarrítmicos	52206.0	0.743
Exames endoscópicos	53692.0	0.745
Número de procedimentos em até 1 ano	88874.5	0.751
Número de procedimentos em até 60 dias	88792.0	0.763

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

Variable	Statistic	p-value
Quantidade de exames histopatológicos	53580.0	0.766
Transfusão de hemoderivados	53572.0	0.768
Tempo entre o P1 e P2	10587.5	0.768
Número de procedimentos em até 30 dias	88715.0	0.776
Teste de esforço	53516.0	0.779
Marca-passo temporário	50100.0	0.785
Intervenção coronária percutânea	53464.0	0.791
Cavografia	53400.0	0.805
Betabloqueador	48664.5	0.807
Citologias	53344.0	0.819
Antiplaquetario EV	50272.0	0.831
PET-CT	53272.0	0.839
Biopsias	53264.0	0.842
Intervenção cardiovascular em laboratório de hemodinâmica	53248.0	0.847
Espirometria / Ergoespirometria	53228.0	0.853
Drenagem de tórax e punção pericárdica ou pleural	53224.0	0.854
Antiviral	50192.0	0.857
Diálise durante a admissão T0	88291.5	0.868
Suporte cardiocirculatório	53156.0	0.878
Ventilação não invasiva	53156.0	0.878
Angio RM	53136.0	0.886
Cirurgia Toracica	53132.0	0.888
Tilt Test	53124.0	0.891
Angiografia	53104.0	0.901
Transplante cardíaco	53088.0	0.909
Angioplastia	53072.0	0.917
Polissonografia	53072.0	0.917
Traqueostomia	53060.0	0.925
Trombolítico	50032.0	0.934
Antiretroviral	50024.0	0.94
Arteriografia	53032.0	0.945
Dieta parenteral	49668.0	0.947
Insuficiência cardíaca	49630.0	0.963
Quantidade de antimicrobianos	50426.0	0.965
Vasodilator	50339.0	0.966
Antibióticos	50333.0	0.972
Stent	53004.0	0.983
Antiplaquetario VO	49988.0	NaN
Hormonio tireoidiano	49988.0	NaN
Broncodilator	49988.0	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
Óbito intraoperatório 1	10194.54	< 0.001
Óbito intraoperatório 3	5824.81	< 0.001
Desfecho principal da admissão T0	267.33	< 0.001
Desfecho final do estudo	52.03	< 0.001
Causa do óbito	1502.67	< 0.001
Tipo de Reoperação 3	40.00	0.001
Classe funcional de IC	134.63	0.002
Tipo de Dispositivo ao final do procedimento 3	55.08	0.003
Raça	34.12	0.005
Insuficiência renal crônica	15.24	0.009
Mudança do tipo de DCEI: entre o Procedimento 1 e Procedimento 2	17.49	0.014
Hemodiálise	64.36	0.015
Tipo de Dispositivo ao final do procedimento 2	19.50	0.019
Insuficiência cardíaca	3.56	0.071
Número de procedimentos	18.82	0.072
Mudança do tipo de DCEI: entre o Procedimento 2 e Procedimento 3	10.74	0.075
Hipertensão arterial	3.52	0.084
Tipo de Dispositivo ao final do procedimento 1	5.02	0.149
Tipo de Reoperação 1	3.65	0.173
Tipo de Procedimento 1	2.39	0.195
Ventilação mecânica / IOT	2.29	0.221
Acidente Vascular Cerebral/ Acidente isquêmico transitório prévios	1.27	0.307
Escolaridade	6.01	0.358
Infarto do miocárdio prévio / Doença arterial coronariana	1.05	0.629
Parada cardíaca prévia/ Taquicardia ventricular instável	0.40	0.636
Admissão em até 180 dias antes da T0	0.82	0.637
Valvopatias/ Prótese valvares	0.79	0.646
Doença cardíaca	4.68	0.704
UTI durante os episódios de hospitalização	0.24	0.711
Tipo de Reoperação 2	1.09	0.727
Óbito intraoperatório 2	0.18	0.752
Doença cardíaca	0.40	0.929
Sexo	0.01	> 0.999
Estado de residência	1.00	> 0.999
Fibrilação / flutter atrial	0.08	> 0.999
Transplante cardíaco prévio	0.01	> 0.999
Endocardite prévia	0.10	> 0.999
Diabetes mellitus	0.10	> 0.999

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

Variable	Statistic	p-value
Doença pulmonar obstrutiva crônica	0.15	> 0.999
Neoplasia em tratamento ou tratada recentemente	0.08	> 0.999
Tipo de Reoperação 4	0.34	> 0.999
Tipo de Dispositivo ao final do procedimento 4	0.34	> 0.999
Óbito intraoperatório 4	0.34	> 0.999
Tipo de Reoperação 5	0.13	> 0.999
Tipo de Dispositivo ao final do procedimento 5	0.13	> 0.999
Óbito intraoperatório 5	0.13	> 0.999
Tipo de Reoperação 6	0.05	> 0.999
Tipo de Dispositivo ao final do procedimento 6	0.06	> 0.999
Óbito intraoperatório 6	0.06	> 0.999
Tipo de Reoperação 7	0.02	> 0.999
Tipo de Dispositivo ao final do procedimento 7	0.02	> 0.999
Óbito intraoperatório 7	0.02	> 0.999
Tipo de Reoperação 8	0.01	> 0.999
Tipo de Dispositivo ao final do procedimento 8	0.01	> 0.999
Óbito intraoperatório 8	0.01	> 0.999
Tipo de Reoperação 9	0.00	> 0.999
Tipo de Dispositivo ao final do procedimento 9	0.00	> 0.999
Óbito intraoperatório 9	0.00	> 0.999
Tipo de Reoperação 10	0.00	> 0.999
Tipo de Dispositivo ao final do procedimento 10	0.00	> 0.999
Óbito intraoperatório 10	0.00	> 0.999
Mudança do tipo de DCEI: entre o Procedimento 3 e Procedimento 4	0.34	> 0.999
Mudança do tipo de DCEI: entre o Procedimento 4 e Procedimento 5	0.13	> 0.999
Mudança do tipo de DCEI: entre o Procedimento 5 e Procedimento 6	0.06	> 0.999
Mudança do tipo de DCEI: entre o Procedimento 6 e Procedimento 7	0.02	> 0.999
Mudança do tipo de DCEI: entre o Procedimento 7 e Procedimento 8	0.01	> 0.999
Mudança do tipo de DCEI: entre o Procedimento 8 e Procedimento 9	0.00	> 0.999
Mudança do tipo de DCEI: entre o Procedimento 9 e Procedimento 10	0.00	> 0.999
Diálise durante os episódios de hospitalização	0.04	> 0.999
Readmissões pós-T0 com diárias de UTI	0.35	> 0.999
Readmissões pós-T0 com diálise	0.01	> 0.999
Readmissão cirúrgica em até 30 dias	0.10	> 0.999
Readmissão cirúrgica entre 31 a 60 dias	0.06	> 0.999
Readmissão cirúrgica entre 61 a 180 dias	0.10	> 0.999
Readmissão cirúrgica em até 1 ano	0.09	> 0.999