

Correlations

Eduardo Yuki Yada

Imports

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

Loading data

```
# df <- readRDS('../dataset/processed_data.rds')
# df_names <- readRDS('../dataset/processed_dictionary.rds')

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){
  df %>%
    kbl(booktabs = T, longtable = T, caption = caption, digits = digits, format = "latex") %>%
    kable_styling(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

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

eligible_columns <- intersect(na_eligible_columns,
                             unique_eligible_columns)

eligible_columns <- setdiff(eligible_columns, weird_columns)
```

```

# df %>% group_by(dieta_enteral) %>% summarise(n = n())
# df %>% group_by(dieta_parenteral) %>% summarise(n = n())

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

  drop_na %>%
  cor %>%
  as.matrix

## Warning in cor(.): o desvio padrão é 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 > 0.8) %>%
  rename_column('row') %>%
  rename_column('column') %>%
  select(row, column, correlation) %>%
  niceFormatting(caption = "Pearson Correlation")

```

Table 1: Pearson Correlation

row	column	correlation
Idade no momento do primeiro procedimento	Idade no Procedimento 1	1.00
Idade no momento do primeiro procedimento	Idade no Procedimento 2	0.99
Ano do procedimento 1	Ano da admissão T0	1.00
Idade no Procedimento 1	Idade no momento do primeiro procedimento	1.00
Idade no Procedimento 1	Idade no Procedimento 2	0.99
Idade no Procedimento 2	Idade no momento do primeiro procedimento	0.99
Idade no Procedimento 2	Idade no Procedimento 1	0.99
Número de atendimentos	Núm. de hospitalizações pós-procedimento	0.85
Núm. de hospitalizações pós-procedimento	Número de atendimentos	0.85
Ano da admissão T0	Ano do procedimento 1	1.00
Readmissão entre 61 a 180 dias	Readmissão em até 1 ano	0.84
Readmissão em até 1 ano	Readmissão entre 61 a 180 dias	0.84
DVA	Diuretico	0.81
Diuretico	DVA	0.81
Vasodilator	Antiviral	0.85
Antiviral	Vasodilator	0.85
Suporte cardiocirculatório	Quantidade de procedimentos invasivos	0.92
Quantidade de procedimentos invasivos	Suporte cardiocirculatório	0.92
Equipe Multiprofissional	Radiografias	0.81
Equipe Multiprofissional	Quantidade de exames diagnóstico por imagem	0.80
ECG	Quantidade de exames por métodos gráficos	1.00
ECG	Exames laboratoriais	0.82

Table 1: Pearson Correlation (*continued*)

row	column	correlation
ECG	Quantidade de exames de análises clínicas	0.82
ECG	Radiografias	0.82
ECG	Quantidade de exames diagnóstico por imagem	0.84
Quantidade de exames por métodos gráficos	ECG	1.00
Quantidade de exames por métodos gráficos	Exames laboratoriais	0.82
Quantidade de exames por métodos gráficos	Quantidade de exames de análises clínicas	0.82
Quantidade de exames por métodos gráficos	Radiografias	0.81
Quantidade de exames por métodos gráficos	Quantidade de exames diagnóstico por imagem	0.83
Exames laboratoriais	ECG	0.82
Exames laboratoriais	Quantidade de exames por métodos gráficos	0.82
Exames laboratoriais	Quantidade de exames de análises clínicas	1.00
Exames laboratoriais	Radiografias	0.82
Exames laboratoriais	Quantidade de exames diagnóstico por imagem	0.87
Quantidade de exames de análises clínicas	ECG	0.82
Quantidade de exames de análises clínicas	Quantidade de exames por métodos gráficos	0.82
Quantidade de exames de análises clínicas	Exames laboratoriais	1.00
Quantidade de exames de análises clínicas	Radiografias	0.82
Quantidade de exames de análises clínicas	Quantidade de exames diagnóstico por imagem	0.87
Biopsias	Quantidade de exames histopatológicos	0.96
Quantidade de exames histopatológicos	Biopsias	0.96
Radiografias	Equipe Multiprofissional	0.81
Radiografias	ECG	0.82
Radiografias	Quantidade de exames por métodos gráficos	0.81
Radiografias	Exames laboratoriais	0.82
Radiografias	Quantidade de exames de análises clínicas	0.82
Radiografias	Quantidade de exames diagnóstico por imagem	0.98
Quantidade de exames diagnóstico por imagem	Equipe Multiprofissional	0.80
Quantidade de exames diagnóstico por imagem	ECG	0.84
Quantidade de exames diagnóstico por imagem	Quantidade de exames por métodos gráficos	0.83
Quantidade de exames diagnóstico por imagem	Exames laboratoriais	0.87
Quantidade de exames diagnóstico por imagem	Quantidade de exames de análises clínicas	0.87
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	3866230.5	< 0.001
Núm. de hospitalizações pós-procedimento	3252935.0	< 0.001
Readmissão em até 30 dias	6482164.0	< 0.001
Readmissão entre 31 a 60 dias	4265729.0	< 0.001
Readmissão entre 61 a 180 dias	0.0	< 0.001
Readmissão em até 1 ano	391988.5	< 0.001
Óbito em até 180 dias após a alta T0	9222357.0	< 0.001
Óbito durante algum episódio de readmissão hospitalar	8400186.0	< 0.001
Tempo entre o P1 e P2	2163785.0	< 0.001
Óbito	8658428.5	< 0.001
Óbito em até 1 ano após a alta T0	8619707.0	< 0.001
Número da Admissão T0	8789301.0	< 0.001
Antiarrítmicos	9334525.5	< 0.001
Transplante cardíaco	10707289.0	< 0.001
Óbito em até 30 dias após a alta T0	10418887.5	< 0.001
DVA	9240747.0	< 0.001
Quantidade de exames diagnóstico por imagem	8622342.0	< 0.001
Antagonista da Aldosterona	9402874.0	< 0.001
Insuficiência cardíaca	9331144.5	< 0.001
Biopsias	10627718.5	< 0.001
Equipe Multiprofissional	9003019.5	< 0.001
Ultrassom	9902163.5	< 0.001
Radiografias	8841677.5	< 0.001
Diurético	9129054.0	< 0.001
Exames laboratoriais	8916723.0	< 0.001
Quantidade de exames de análises clínicas	8917747.0	< 0.001
Núm. de hospitalizações pré-procedimento	9245314.0	< 0.001
Quantidade de exames por métodos gráficos	8888305.5	< 0.001
ECG	8942458.5	< 0.001
Ecocardiograma	9331342.0	< 0.001
Quantidade de classes medicamentosas utilizadas	8969091.5	< 0.001
Quantidade de procedimentos invasivos	9691531.0	< 0.001
Quantidade de exames histopatológicos	10566514.0	< 0.001
Culturas	9880723.0	< 0.001
Tempo de seguimento total	12688568.0	< 0.001
Cateterismo	10044385.5	< 0.001

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

Variable	Statistic	p-value
Número de comorbidades	9232257.5	< 0.001
Anticoagulantes orais	10312854.0	< 0.001
Antiviral	10725929.0	< 0.001
Ano do procedimento 2	1582533.0	< 0.001
Cateter venoso central	10437573.5	< 0.001
Holter	10151027.5	< 0.001
Ressonancia magnetica	10284346.5	< 0.001
Vasodilator	9722667.0	< 0.001
Tomografia	10138767.5	< 0.001
Cintilografia	10437555.0	< 0.001
Tempo de sobrevida	467277.0	< 0.001
Psicofármacos	9638523.5	< 0.001
Bloqueador do canal de calcio	10590134.0	< 0.001
Exames endoscópicos	10641489.0	< 0.001
Digoxina	10411465.0	< 0.001
Óbito em até 2 anos após a alta T0	8230328.0	< 0.001
Estatinas	9849016.0	< 0.001
Antifúngicos	10600974.5	< 0.001
Betabloqueador	10306552.5	< 0.001
Suporte cardiocirculatório	10782603.5	< 0.001
Antibióticos	9757970.0	< 0.001
Ano do procedimento 3	193655.5	< 0.001
Instalação de CEC	10688515.0	< 0.001
IECA/BRA	9870335.0	< 0.001
Eletrofisiologia	10560665.5	< 0.001
Diárias no serviço de Emergência na admissão T0	10436103.5	< 0.001
Outros procedimentos cirúrgicos	10504086.5	< 0.001
Tempo entre o P2 e P3	185266.0	< 0.001
Idade no Procedimento 2	1447376.0	< 0.001
Antiplaquetario EV	10786836.5	< 0.001
Óbito hospitalar	11089602.0	< 0.001
Bomba de infusão contínua	10711981.5	< 0.001
Insulina	10576673.0	< 0.001
Citologias	10790037.0	< 0.001
Idade no momento do primeiro procedimento	11654245.0	< 0.001
Idade no Procedimento 1	11654245.0	< 0.001
Intervenção coronária percutânea	10787637.0	< 0.001
Óbito em até 3 anos após a alta T0	7907761.0	< 0.001
Transfusão de hemoderivados	10778171.0	< 0.001
Cardioversão/ Desfibrilação	10826535.0	< 0.001
Anticonvulsivante	10702354.0	< 0.001
Broncodilator	10742016.0	< 0.001
Angio RM	10846563.0	< 0.001
Interconsulta médica	10596486.5	< 0.001
Espirometria / Ergoespirometria	10839768.5	0.002
Angio TC	10766824.0	0.004
Tilt Test	10860355.5	0.011
Antiretroviral	10875930.0	0.013
Teste de esforço	10829520.5	0.016
Flebografia	10809071.5	0.017
Hormonio tireoidiano	10814428.5	0.018
PET-CT	10848740.5	0.021
Intervenção cardiovascular em laboratório de hemodinâmica	10851586.5	0.023

Table 2: Mann-Whitney Test (continued)

Variable	Statistic	p-value
Ventilação não invasiva	10922463.5	0.045
Número de Mudanças do tipo de DCEI	1238816.5	0.054
Angioplastia	10874662.0	0.061
Marca-passo temporário	10851907.5	0.099
Arteriografia	10883193.0	0.128
Dieta enteral	10877137.5	0.222
Cirurgia Toracica	10877860.0	0.251
Antihipertensivo	10835004.5	0.268
Drenagem de tórax e punção pericárdica ou pleural	10912644.5	0.274
Polissonografia	10882670.0	0.286
Trombolítico	10885448.5	0.314
Ano da admissão T0	11020537.5	0.322
Aortografia	10881396.0	0.329
Ano do procedimento 1	11053133.0	0.348
Idade no Procedimento 3	157240.5	0.373
Dieta parenteral	10898479.5	0.395
Hipoglicemiante	10854525.5	0.472
Cavografia	10880001.5	0.57
Cirurgia Cardiovascular	10919315.0	0.604
Antiplaquetario VO	10916188.0	0.626
Angiografia	10888672.0	0.701
Traqueostomia	10896461.0	0.721
Stent	10893982.5	0.748

```
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	39.23	< 0.001
Doença cardíaca	69.33	< 0.001
Doença cardíaca	34.98	< 0.001
Classe funcional de IC	50.08	< 0.001
Infarto do miocárdio prévio / Doença arterial coronariana	31.11	< 0.001
Insuficiência cardíaca	143.38	< 0.001
Fibrilação / flutter atrial	16.71	< 0.001
Parada cardíaca prévia/ Taquicardia ventricular instável	21.25	< 0.001
Transplante cardíaco prévio	88.94	< 0.001
Valvopatias/ Prótese valvares	15.48	< 0.001
Número de procedimentos	155.14	< 0.001
Tipo de Procedimento 1	115.44	< 0.001
Tipo de Reoperação 1	138.59	< 0.001
Tipo de Dispositivo ao final do procedimento 1	221.25	< 0.001
Tipo de Reoperação 2	884.11	< 0.001
Tipo de Dispositivo ao final do procedimento 2	116.20	< 0.001
Óbito intraoperatório 2	65.95	< 0.001
Tipo de Reoperação 3	90.17	< 0.001
Tipo de Dispositivo ao final do procedimento 3	106.23	< 0.001
Óbito intraoperatório 3	103.05	< 0.001
Tipo de Reoperação 4	103.78	< 0.001
Tipo de Dispositivo ao final do procedimento 4	102.86	< 0.001
Óbito intraoperatório 4	95.36	< 0.001
Mudança do tipo de DCEI: entre o Procedimento 1 e Procedimento 2	64.50	< 0.001
Mudança do tipo de DCEI: entre o Procedimento 2 e Procedimento 3	102.72	< 0.001
Mudança do tipo de DCEI: entre o Procedimento 3 e Procedimento 4	95.98	< 0.001
UTI durante os episódios de hospitalização	228.35	< 0.001
Admissão em até 180 dias antes da T0	173.36	< 0.001
UTI durante a admissão T0	2695.18	< 0.001
Readmissões pós-T0 com diárias de UTI	346.95	< 0.001
Desfecho principal da admissão T0	27.05	< 0.001
Desfecho final do estudo	340.69	< 0.001
Causa do óbito	83.33	< 0.001
Ventilação mecânica / IOT	81.72	< 0.001
Sexo	10.77	< 0.001
Óbito intraoperatório 5	14.09	< 0.001
Óbito intraoperatório 6	12.60	< 0.001
Diálise durante os episódios de hospitalização	15.60	< 0.001
Diabetes mellitus	12.15	0.001
Tipo de Reoperação 5	18.92	0.001
Mudança do tipo de DCEI: entre o Procedimento 4 e Procedimento 5	17.66	0.001
Tipo de Dispositivo ao final do procedimento 5	19.50	0.003
Mudança do tipo de DCEI: entre o Procedimento 5 e Procedimento 6	16.48	0.003
Tipo de Reoperação 6	16.21	0.004
Insuficiência renal crônica	7.41	0.008
Diálise durante a admissão T0	55.02	0.008
Hemodiálise	8.35	0.011
Estado de residência	52.87	0.013
Tipo de Dispositivo ao final do procedimento 6	14.68	0.02
Tipo de Dispositivo ao final do procedimento 7	14.99	0.021
Tipo de Reoperação 7	11.34	0.025
Óbito intraoperatório 7	5.93	0.025
Mudança do tipo de DCEI: entre o Procedimento 6 e Procedimento 7	6.87	0.041

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

Variable	Statistic	p-value
Endocardite prévia	4.34	0.045
Readmissões pós-T0 com diálise	10.84	0.064
Doença pulmonar obstrutiva crônica	3.09	0.078
Neoplasia em tratamento ou tratada recentemente	2.72	0.103
Acidente Vascular Cerebral/ Acidente isquêmico transitório prévios	1.74	0.19
Óbito intraoperatório	1.14	0.406
Hipertensão arterial	0.52	0.493
Óbito intraoperatório 8	0.76	0.608
Raça	4.22	0.616
Tipo de Dispositivo ao final do procedimento 8	0.76	0.621
Tipo de Reoperação 8	0.76	0.625
Óbito intraoperatório 1	0.72	0.626
Mudança do tipo de DCEI: entre o Procedimento 7 e Procedimento 8	1.12	0.634
Tipo de Reoperação 9	0.52	> 0.999
Tipo de Dispositivo ao final do procedimento 9	0.52	> 0.999
Óbito intraoperatório 9	0.52	> 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	0.52	> 0.999
Mudança do tipo de DCEI: entre o Procedimento 9 e Procedimento 10	0.10	> 0.999