

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))
}
```

```

}

## Can't calculate Wilcox test for variable death_3year

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 durante algum episódio de readmissão hospitalar	301957.0	< 0.001
Óbito em até 2 anos após a alta T0	0.0	< 0.001
Óbito	211285.5	< 0.001
Tempo de sobrevida	233887.0	< 0.001
Núm. de hospitalizações pós-procedimento	789555.0	< 0.001
Número de atendimentos	807195.0	< 0.001
Readmissão em até 1 ano	1223047.5	< 0.001
Tempo de seguimento total	2780838.0	< 0.001
Número da Admissão T0	1219318.5	< 0.001
Antagonista da Aldosterona	1382408.0	< 0.001
Núm. de hospitalizações pré-procedimento	1303677.5	< 0.001
Insuficiência cardíaca	1381802.0	< 0.001
Readmissão entre 61 a 180 dias	1526058.5	< 0.001
Antiarrítmicos	1516404.5	< 0.001
Diuretico	1412882.5	< 0.001
Tempo entre o P1 e P2	88028.5	< 0.001
DVA	1518986.0	< 0.001
Número de Mudanças do tipo de DCEI	32938.5	< 0.001
Digoxina	1660584.5	< 0.001
Quantidade de classes medicamentosas utilizadas	1436748.0	< 0.001
Anticoagulantes orais	1665692.0	< 0.001
Exames laboratoriais	1453111.5	< 0.001
Quantidade de exames de análises clínicas	1453612.0	< 0.001
Cintilografia	1686126.5	< 0.001
Quantidade de exames diagnóstico por imagem	1468482.5	< 0.001
Quantidade de exames por métodos gráficos	1478928.5	< 0.001
ECCG	1483137.5	< 0.001
Estatinas	1543002.0	< 0.001
Readmissão em até 30 dias	1690889.0	< 0.001
Equipe Multiprofissional	1537728.5	< 0.001
Ultrassom	1662807.0	< 0.001
Ressonancia magnetica	1677681.0	< 0.001
Radiografias	1519889.5	< 0.001
Cateterismo	1653175.5	< 0.001
Readmissão entre 31 a 60 dias	1681905.0	< 0.001

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

Variable	Statistic	p-value
Psicofármacos	1575805.5	< 0.001
Angioplastia	1766157.5	< 0.001
Ecocardiograma	1594087.5	< 0.001
Bomba de infusão contínua	1734474.5	< 0.001
Vasodilator	1618300.0	0.001
Quantidade de procedimentos invasivos	1647799.0	0.001
Culturas	1667122.5	0.002
Holter	1685158.5	0.002
IECA/BRA	1596489.5	0.003
Tempo entre o P2 e P3	6294.0	0.005
Antibióticos	1596641.0	0.005
Intervenção coronária percutânea	1753296.5	0.006
Transfusão de hemoderivados	1754321.5	0.01
Número de comorbidades	1624311.0	0.016
Antiplaquetario EV	1763478.5	0.025
Intervenção cardiovascular em laboratório de hemodinâmica	1763664.0	0.028
Ano do procedimento 2	61189.5	0.034
Flebografia	1750973.5	0.035
Eletrofisiologia	1737172.0	0.037
Idade no Procedimento 2	63488.0	0.04
Tomografia	1720132.5	0.051
Ano da admissão T0	1908140.5	0.051
Ano do procedimento 1	1908571.0	0.055
Diárias no serviço de Emergência na admissão T0	1729756.0	0.076
Polissonografia	1773666.5	0.109
Insulina	1742141.5	0.132
Exames endoscópicos	1764096.5	0.147
Outros procedimentos cirúrgicos	1745811.5	0.18
Angio TC	1757342.0	0.189
Antiplaquetario VO	1801886.5	0.22
Cavografia	1789942.5	0.237
Dieta enteral	1774500.0	0.257
Hipoglicemiante	1803409.5	0.261
Instalação de CEC	1767378.0	0.299
Biopsias	1787335.5	0.306
Interconsulta médica	1813244.5	0.325
Broncodilator	1765239.0	0.349
Antihipertensivo	1797928.0	0.367
Drenagem de tórax e punção pericárdica ou pleural	1785202.5	0.383
Tilt Test	1775206.0	0.388
Cardioversão/ Desfibrilação	1784847.0	0.399
Anticonvulsivante	1765307.5	0.443
Bloqueador do canal de calcio	1789528.5	0.465
Teste de esforço	1787177.0	0.473
Angio RM	1783306.5	0.477
Suporte cardiocirculatório	1775785.0	0.491
Idade no Procedimento 3	3000.5	0.496
Transplante cardíaco	1782832.5	0.506
Ventilação não invasiva	1782595.5	0.522
Hormonio tireoidiano	1771378.5	0.528
Angiografia	1782240.0	0.547
Cirurgia Toracica	1782240.0	0.547
Aortografia	1782121.5	0.556

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

Variable	Statistic	p-value
Cateter venoso central	1769784.0	0.586
Quantidade de exames histopatológicos	1784477.5	0.608
Antifúngicos	1773016.5	0.626
Traqueostomia	1781055.0	0.651
Antiviral	1776999.0	0.684
Idade no momento do primeiro procedimento	1754354.5	0.708
Idade no Procedimento 1	1754354.5	0.708
Citologias	1777222.5	0.717
Trombolítico	1780462.5	0.722
Antiretroviral	1780462.5	0.722
Arteriografia	1780225.5	0.758
Dieta parenteral	1780107.0	0.779
Marca-passo temporário	1777181.5	0.812
Ano do procedimento 3	3453.0	0.841
Betabloqueador	1785307.5	0.856
Espirometria / Ergoespirometria	1778287.5	0.859
PET-CT	1778525.0	0.888
Cirurgia Cardiovascular	1777216.0	0.906
Óbito hospitalar	1779514.5	NaN
Óbito em até 30 dias após a alta T0	1779514.5	NaN
Óbito em até 180 dias após a alta T0	1779514.5	NaN
Óbito em até 1 ano após a alta T0	1779514.5	NaN
Óbito em até 3 anos após a alta T0	NaN	NaN
Stent	1779514.5	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))
  }
}

## Can't calculate Chi Squared test for variable death_intraop_1
## Can't calculate Chi Squared test for variable disch_outcomes_t0
## Can't calculate Chi Squared test for variable death_intraop

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	40.84	< 0.001
Insuficiência cardíaca	45.13	< 0.001
Tipo de Dispositivo ao final do procedimento 1	66.07	< 0.001
Tipo de Reoperação 2	65.46	< 0.001
Tipo de Dispositivo ao final do procedimento 2	75.38	< 0.001
Óbito intraoperatório 2	57.61	< 0.001
Mudança do tipo de DCEI: entre o Procedimento 1 e Procedimento 2	72.72	< 0.001
UTI durante os episódios de hospitalização	35.95	< 0.001
Admissão em até 180 dias antes da T0	49.31	< 0.001
UTI durante a admissão T0	2694.67	< 0.001
Desfecho final do estudo	1577.21	< 0.001
Causa do óbito	330.49	< 0.001
Ventilação mecânica / IOT	11.63	< 0.001
Óbito intraoperatório 3	15.52	< 0.001
Classe funcional de IC	35.35	0.001
Mudança do tipo de DCEI: entre o Procedimento 2 e Procedimento 3	15.58	0.001
Tipo de Procedimento 1	10.49	0.002
Tipo de Dispositivo ao final do procedimento 3	17.59	0.005
Óbito intraoperatório 4	7.89	0.006
Tipo de Reoperação 3	15.16	0.009
Número de procedimentos	58.31	0.015
Tipo de Reoperação 1	11.25	0.018
Doença cardíaca	22.69	0.021
Mudança do tipo de DCEI: entre o Procedimento 3 e Procedimento 4	7.89	0.026
Hipertensão arterial	4.48	0.04
Readmissões pós-T0 com diálise	30.94	0.051
Tipo de Reoperação 4	7.74	0.053
Readmissões pós-T0 com diárias de UTI	20.06	0.077
Sexo	2.75	0.102
Diálise durante os episódios de hospitalização	3.78	0.103
Estado de residência	48.74	0.103
Tipo de Dispositivo ao final do procedimento 4	7.89	0.103
Raça	9.65	0.115
Diálise durante a admissão T0	63.63	0.117
Óbito intraoperatório 5	3.07	0.126
Insuficiência renal crônica	2.24	0.147
Valvopatias/ Prótese valvares	2.53	0.148
Fibrilação / flutter atrial	1.59	0.234
Mudança do tipo de DCEI: entre o Procedimento 4 e Procedimento 5	3.05	0.24
Doença cardíaca	4.37	0.245
Acidente Vascular Cerebral/ Acidente isquêmico transitório prévios	1.06	0.347
Tipo de Reoperação 5	3.05	0.355
Óbito intraoperatório 6	1.30	0.421
Tipo de Dispositivo ao final do procedimento 5	3.05	0.458
Mudança do tipo de DCEI: entre o Procedimento 5 e Procedimento 6	1.30	0.468
Infarto do miocárdio prévio / Doença arterial coronariana	0.45	0.557
Doença pulmonar obstrutiva crônica	0.37	0.608
Transplante cardíaco prévio	0.65	0.642
Tipo de Reoperação 6	1.25	0.703

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

Variable	Statistic	p-value
Endocardite prévia	0.49	0.719
Neoplasia em tratamento ou tratada recentemente	0.26	0.743
Parada cardíaca prévia/ Taquicardia ventricular instável	0.14	0.769
Tipo de Dispositivo ao final do procedimento 6	1.30	0.831
Diabetes melittus	0.08	0.852
Hemodiálise	0.20	> 0.999
Tipo de Reoperação 7	0.51	> 0.999
Tipo de Dispositivo ao final do procedimento 7	0.49	> 0.999
Óbito intraoperatório 7	0.51	> 0.999
Tipo de Reoperação 8	0.19	> 0.999
Tipo de Dispositivo ao final do procedimento 8	0.19	> 0.999
Óbito intraoperatório 8	0.19	> 0.999
Tipo de Reoperação 9	0.08	> 0.999
Tipo de Dispositivo ao final do procedimento 9	0.08	> 0.999
Óbito intraoperatório 9	0.08	> 0.999
Tipo de Reoperação 10	0.02	> 0.999
Tipo de Dispositivo ao final do procedimento 10	0.02	> 0.999
Óbito intraoperatório 10	0.02	> 0.999
Mudança do tipo de DCEI: entre o Procedimento 6 e Procedimento 7	0.49	> 0.999
Mudança do tipo de DCEI: entre o Procedimento 7 e Procedimento 8	0.19	> 0.999
Mudança do tipo de DCEI: entre o Procedimento 8 e Procedimento 9	0.08	> 0.999
Mudança do tipo de DCEI: entre o Procedimento 9 e Procedimento 10	0.02	> 0.999
Óbito intraoperatório 1	NaN	NaN
Desfecho principal da admissão T0	NaN	NaN
Óbito intraoperatório	NaN	NaN