

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_1year
## Can't calculate Wilcox test for variable death_2year
## 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 em até 180 dias após a alta T0	0.0	< 0.001
Óbito	285806.5	< 0.001
Óbito durante algum episódio de readmissão hospitalar	914759.0	< 0.001
Readmissão entre 61 a 180 dias	845941.0	< 0.001
Readmissão entre 31 a 60 dias	1193824.5	< 0.001
Readmissão em até 1 ano	913669.5	< 0.001
Tempo de sobrevida	365106.5	< 0.001
Tempo de seguimento total	3658610.5	< 0.001
Readmissão em até 30 dias	1523277.0	< 0.001
Antagonista da Aldosterona	1498453.5	< 0.001
Número de comorbidades	1300415.5	< 0.001
Insuficiência cardíaca	1503076.0	< 0.001
Diuretico	1388195.0	< 0.001
DVA	1526061.5	< 0.001
Culturas	1626890.5	< 0.001
Quantidade de classes medicamentosas utilizadas	1323142.5	< 0.001
Exames laboratoriais	1350697.5	< 0.001
Quantidade de exames de análises clínicas	1350763.0	< 0.001
Ultrassom	1715570.0	< 0.001
Equipe Multiprofissional	1473255.5	< 0.001
Antiarrítmicos	1637562.0	< 0.001
Vasodilator	1555256.5	< 0.001
Quantidade de exames diagnóstico por imagem	1413568.5	< 0.001
ECG	1419855.0	< 0.001
Quantidade de exames por métodos gráficos	1422075.5	< 0.001
Radiografias	1466908.5	< 0.001
Anticoagulantes orais	1815219.0	< 0.001
Tomografia	1754159.0	< 0.001
Tempo entre o P1 e P2	99266.0	< 0.001
Insulina	1809099.0	< 0.001
Número da Admissão T0	1628979.5	< 0.001
Ecocardiograma	1611734.0	< 0.001

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

Variable	Statistic	p-value
Estatinas	1618424.5	< 0.001
Citologias	1951885.0	< 0.001
Psicofármacos	1625660.5	< 0.001
Número de atendimentos	1612342.0	< 0.001
Núm. de hospitalizações pré-procedimento	1674014.0	< 0.001
Antibióticos	1611441.0	< 0.001
Cintilografia	1880510.0	< 0.001
Núm. de hospitalizações pós-procedimento	1646191.0	< 0.001
Ressonancia magnetica	1860538.5	< 0.001
Ano da admissão T0	1628067.0	< 0.001
Ano do procedimento 1	1637224.0	< 0.001
Antiplaquetario EV	1960826.0	< 0.001
Interconsulta médica	1832267.5	< 0.001
Quantidade de exames histopatológicos	1952150.5	< 0.001
Idade no momento do primeiro procedimento	1687273.5	< 0.001
Idade no Procedimento 1	1687273.5	< 0.001
Quantidade de procedimentos invasivos	1810284.5	< 0.001
Ano do procedimento 2	80760.5	< 0.001
IECA/BRA	1723684.0	< 0.001
Diárias no serviço de Emergência na admissão T0	1874635.5	< 0.001
Aortografia	1979289.0	< 0.001
Intervenção coronária percutânea	1958696.0	< 0.001
Holter	1871670.0	< 0.001
Cateterismo	1875364.0	< 0.001
Digoxina	1919181.0	0.001
Tempo entre o P2 e P3	7727.5	0.001
Suporte cardiocirculatório	1980706.5	0.002
Transfusão de hemoderivados	1967479.0	0.003
Arteriografia	1992666.5	0.009
Outros procedimentos cirúrgicos	1928157.5	0.009
Antiplaquetario VO	1949689.0	0.012
Antifúngicos	1965095.5	0.018
Cateter venoso central	1955840.0	0.024
Tilt Test	1987925.0	0.036
Ano do procedimento 3	6390.5	0.055
Bomba de infusão contínua	1972561.0	0.063
Exames endoscópicos	1978108.5	0.066
Hipoglicemiante	1960281.0	0.087
Anticonvulsivante	1966029.5	0.093
Hormonio tireoidiano	1977907.5	0.12
Idade no Procedimento 3	2781.5	0.12
Teste de esforço	2016315.0	0.14
Polissonografia	1993833.0	0.149
Angioplastia	1993961.5	0.171
Betabloqueador	1952689.5	0.171
Drenagem de tórax e punção pericárdica ou pleural	1990496.5	0.201
Cavografia	1987973.5	0.224
Intervenção cardiovascular em laboratório de hemodinâmica	2007250.0	0.319
Dieta enteral	1994857.0	0.332
Cirurgia Toracica	1995117.5	0.378
Biopsias	1992732.0	0.403
Ventilação não invasiva	1995383.5	0.424
Transplante cardíaco	1995515.0	0.447

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

Variable	Statistic	p-value
Flebografia	1988607.0	0.454
Angio RM	2003883.0	0.456
Broncodilator	1988062.0	0.482
Eletrofisiologia	1984931.0	0.503
Angiografia	2002717.5	0.526
Número de Mudanças do tipo de DCEI	55101.5	0.528
Idade no Procedimento 2	50001.5	0.63
Traqueostomia	2001293.0	0.64
Instalação de CEC	2004502.0	0.697
Trombolítico	2000645.5	0.714
Antiretroviral	2000645.5	0.714
Antihipertensivo	1993171.5	0.77
Dieta parenteral	2000257.0	0.772
Antiviral	1997733.0	0.782
Angio TC	2004508.5	0.789
Cardioversão/ Desfibrilação	1997855.0	0.798
Bloqueador do canal de calcio	1996843.0	0.851
Cirurgia Cardiovascular	2002113.0	0.904
Marca-passo temporário	1998359.0	0.906
Espirometria / Ergoespirometria	1999142.0	0.951
PET-CT	1999531.0	0.992
Óbito hospitalar	1999609.5	NaN
Óbito em até 30 dias após a alta T0	1999609.5	NaN
Óbito em até 1 ano após a alta T0	NaN	NaN
Óbito em até 2 anos após a alta T0	NaN	NaN
Óbito em até 3 anos após a alta T0	NaN	NaN
Stent	1999609.5	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
Doença cardíaca	24.40	< 0.001
Classe funcional de IC	75.91	< 0.001
Hipertensão arterial	23.47	< 0.001
Infarto do miocárdio prévio / Doença arterial coronariana	35.43	< 0.001
Insuficiência cardíaca	60.46	< 0.001
Fibrilação / flutter atrial	20.65	< 0.001
Valvopatias/ Prótese valvares	32.81	< 0.001
Diabetes mellitus	49.58	< 0.001
Insuficiência renal crônica	50.43	< 0.001
Hemodiálise	47.01	< 0.001
Acidente Vascular Cerebral/ Acidente isquêmico transitório prévios	28.06	< 0.001
Tipo de Dispositivo ao final do procedimento 1	62.54	< 0.001
Tipo de Reoperação 2	72.44	< 0.001
Tipo de Dispositivo ao final do procedimento 2	65.17	< 0.001
Óbito intraoperatório 2	61.99	< 0.001
Óbito intraoperatório 3	15.45	< 0.001
Mudança do tipo de DCEI: entre o Procedimento 1 e Procedimento 2	61.84	< 0.001
UTI durante os episódios de hospitalização	33.76	< 0.001
Admissão em até 180 dias antes da T0	37.64	< 0.001
UTI durante a admissão T0	2418.82	< 0.001
Desfecho final do estudo	1413.26	< 0.001
Causa do óbito	1928.92	< 0.001
Tipo de Procedimento 1	14.46	< 0.001
Diálise durante os episódios de hospitalização	36.91	< 0.001
Ventilação mecânica / IOT	14.01	< 0.001
Mudança do tipo de DCEI: entre o Procedimento 2 e Procedimento 3	15.53	0.001
Diálise durante a admissão T0	156.75	0.001
Tipo de Reoperação 3	15.55	0.004
Tipo de Reoperação 1	15.08	0.005
Doença cardíaca	26.49	0.009
Escolaridade	17.08	0.01
Doença pulmonar obstrutiva crônica	8.92	0.01
Tipo de Dispositivo ao final do procedimento 3	16.14	0.01
Óbito intraoperatório 4	6.43	0.01
Sexo	5.97	0.015
Número de procedimentos	62.31	0.018
Parada cardíaca prévia/ Taquicardia ventricular instável	5.25	0.022
Neoplasia em tratamento ou tratada recentemente	5.51	0.033
Mudança do tipo de DCEI: entre o Procedimento 3 e Procedimento 4	6.43	0.045
Readmissões pós-T0 com diárias de UTI	26.25	0.072
Óbito intraoperatório 5	3.26	0.077
Tipo de Reoperação 4	6.38	0.082
Readmissões pós-T0 com diálise	9.29	0.088
Tipo de Dispositivo ao final do procedimento 4	6.51	0.148
Mudança do tipo de DCEI: entre o Procedimento 4 e Procedimento 5	3.24	0.215
Raça	6.83	0.258

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

Variable	Statistic	p-value
Tipo de Reoperação 5	3.24	0.311
Óbito intraoperatório 6	1.38	0.398
Tipo de Dispositivo ao final do procedimento 5	3.24	0.429
Mudança do tipo de DCEI: entre o Procedimento 5 e Procedimento 6	1.38	0.469
Endocardite prévia	0.65	0.532
Estado de residência	19.84	0.592
Óbito intraoperatório 7	0.54	0.687
Tipo de Reoperação 6	1.33	0.714
Tipo de Dispositivo ao final do procedimento 6	1.38	0.731
Transplante cardíaco prévio	0.14	> 0.999
Tipo de Reoperação 7	0.54	> 0.999
Tipo de Dispositivo ao final do procedimento 7	0.52	> 0.999
Tipo de Reoperação 8	0.20	> 0.999
Tipo de Dispositivo ao final do procedimento 8	0.20	> 0.999
Óbito intraoperatório 8	0.20	> 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	0.52	> 0.999
Mudança do tipo de DCEI: entre o Procedimento 7 e Procedimento 8	0.20	> 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