

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_30days
## Can't calculate Wilcox test for variable death_180days
## 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 hospitalar	0.0	< 0.001
Antifúngicos	1437332.5	< 0.001
Culturas	939195.0	< 0.001
DVA	743242.5	< 0.001
Tempo de seguimento total	3880257.5	< 0.001
Ultrassom	1269855.0	< 0.001
Tempo de sobrevida	438455.5	< 0.001
Citologias	1852834.0	< 0.001
Ecocardiograma	796209.0	< 0.001
Exames endoscópicos	1776776.0	< 0.001
Tomografia	1318495.5	< 0.001
Equipe Multiprofissional	766720.5	< 0.001
Quantidade de exames histopatológicos	1822138.0	< 0.001
Quantidade de exames diagnóstico por imagem	699518.0	< 0.001
Diuretico	927841.5	< 0.001
Radiografias	798278.5	< 0.001
Anticonvulsivante	1684504.5	< 0.001
Transfusão de hemoderivados	1859947.0	< 0.001
Quantidade de exames de análises clínicas	880578.5	< 0.001
Exames laboratoriais	880702.5	< 0.001
Outros procedimentos cirúrgicos	1590944.0	< 0.001
Quantidade de exames por métodos gráficos	930545.0	< 0.001
ECG	932160.0	< 0.001
Quantidade de procedimentos invasivos	1355207.5	< 0.001
Insulina	1653494.0	< 0.001
Antibióticos	1054166.5	< 0.001
Psicofármacos	1210915.0	< 0.001
Antiarrítmicos	1463467.0	< 0.001
Núm. de hospitalizações pós-procedimento	2900048.5	< 0.001
Quantidade de classes medicamentosas utilizadas	1180455.0	< 0.001

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

Variable	Statistic	p-value
Vasodilator	1401710.5	< 0.001
Broncodilator	1858485.5	< 0.001
Cateter venoso central	1833592.0	< 0.001
Ventilação não invasiva	1998878.0	< 0.001
Suporte cardiocirculatório	1999707.0	< 0.001
Dieta enteral	2014036.5	< 0.001
Cirurgia Toracica	2014204.0	< 0.001
Bomba de infusão contínua	1928054.5	< 0.001
Antagonista da Aldosterona	1678615.5	< 0.001
Instalação de CEC	1960729.0	< 0.001
Stent	2058115.5	< 0.001
Cateterismo	1787115.5	< 0.001
Antiviral	2009085.5	< 0.001
Cirurgia Cardiovascular	1911131.0	< 0.001
Número de comorbidades	1569057.0	< 0.001
Insuficiência cardíaca	1713251.0	< 0.001
Trombolítico	2043388.5	< 0.001
PET-CT	2010890.0	< 0.001
Tempo entre o P1 e P2	102939.5	< 0.001
Angio TC	1944762.0	< 0.001
Drenagem de tórax e punção pericárdica ou pleural	2017438.5	< 0.001
Readmissão em até 1 ano	2330883.0	< 0.001
Número da Admissão T0	1732827.0	< 0.001
Dieta parenteral	2050881.0	< 0.001
Interconsulta médica	1851033.0	< 0.001
Núm. de hospitalizações pré-procedimento	1747301.5	< 0.001
Holter	1881677.0	< 0.001
Readmissão entre 61 a 180 dias	2262370.0	< 0.001
Óbito	2311607.5	< 0.001
Número de atendimentos	2386689.0	< 0.001
Antiplaquetario EV	2027339.5	< 0.001
Diárias no serviço de Emergência na admissão T0	1926734.5	< 0.001
Biopsias	2027734.0	< 0.001
Cavografia	2022985.5	< 0.001
Cintilografia	1972745.5	< 0.001
Readmissão entre 31 a 60 dias	2185473.0	< 0.001
Óbito durante algum episódio de readmissão hospitalar	2204890.0	< 0.001
Intervenção coronária percutânea	2024936.0	< 0.001
Traqueostomia	2051921.5	< 0.001
Readmissão em até 30 dias	2145518.0	0.001
Polissonografia	2052329.0	0.002
Antihipertensivo	1996508.0	0.002
Anticoagulantes orais	1992152.0	0.004
Hipoglicemiante	2130105.5	0.006
Cardioversão/ Desfibrilação	2048457.5	0.014
Arteriografia	2059033.0	0.015
Aortografia	2053508.0	0.018
Antiretroviral	2059167.0	0.025
Flebografia	2031872.5	0.026
Bloqueador do canal de calcio	2031972.5	0.026
Digoxina	2010861.5	0.033
Betabloqueador	1992442.0	0.038
Tempo entre o P2 e P3	7485.5	0.039

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

Variable	Statistic	p-value
Ano do procedimento 2	70593.0	0.073
Hormonio tireoidiano	2044689.5	0.141
Teste de esforço	2082900.0	0.142
Idade no Procedimento 2	68178.0	0.151
Angioplastia	2060334.5	0.189
Estatinas	2000379.0	0.265
Espirometria / Ergoespirometria	2073468.0	0.33
Eletrofisiologia	2044078.0	0.332
Número de Mudanças do tipo de DCEI	53165.5	0.409
Marca-passo temporário	2056995.0	0.409
Ano do procedimento 3	6060.5	0.422
Transplante cardíaco	2061914.5	0.463
Tilt Test	2070062.0	0.473
Angiografia	2069407.0	0.511
Idade no momento do primeiro procedimento	2111765.5	0.538
Idade no Procedimento 1	2111765.5	0.538
Angio RM	2062446.0	0.548
Ano do procedimento 1	2022535.0	0.558
Ano da admissão T0	2024502.5	0.6
Ressonancia magnetica	2078568.5	0.659
IECA/BRA	2090176.5	0.721
Idade no Procedimento 3	4965.5	0.867
Antiplaquetario VO	2064003.0	0.922
Intervenção cardiovascular em laboratório de hemodinâmica	2066120.0	0.988
Óbito em até 30 dias após a alta T0	NaN	NaN
Óbito em até 180 dias após a alta T0	NaN	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

```
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
Raça	185.91	< 0.001
Escolaridade	100.27	< 0.001
Doença cardíaca	84.49	< 0.001
Classe funcional de IC	54.38	< 0.001
Insuficiência cardíaca	117.78	< 0.001
Diabetes mellitus	50.28	< 0.001
Insuficiência renal crônica	40.92	< 0.001
Tipo de Procedimento 1	26.58	< 0.001
Tipo de Reoperação 1	51.41	< 0.001
Tipo de Dispositivo ao final do procedimento 1	37.27	< 0.001
Óbito intraoperatório 1	421.55	< 0.001
Tipo de Reoperação 2	67.99	< 0.001
Tipo de Dispositivo ao final do procedimento 2	60.03	< 0.001
Óbito intraoperatório 2	56.95	< 0.001
Óbito intraoperatório 3	260.84	< 0.001
Mudança do tipo de DCEI: entre o Procedimento 1 e Procedimento 2	59.48	< 0.001
Diálise durante os episódios de hospitalização	252.86	< 0.001
UTI durante os episódios de hospitalização	436.28	< 0.001
Admissão em até 180 dias antes da T0	42.83	< 0.001
UTI durante a admissão T0	9340.99	< 0.001
Diálise durante a admissão T0	702.59	< 0.001
Desfecho principal da admissão T0	15784.22	< 0.001
Desfecho final do estudo	1259.35	< 0.001
Óbito intraoperatório	662.60	< 0.001
Causa do óbito	1265.01	< 0.001
Ventilação mecânica / IOT	192.37	< 0.001
Hemodiálise	19.74	0.004
Mudança do tipo de DCEI: entre o Procedimento 2 e Procedimento 3	13.80	0.007
Óbito intraoperatório 4	8.31	0.007
Tipo de Dispositivo ao final do procedimento 3	16.19	0.008
Doença cardíaca	11.24	0.01
Tipo de Reoperação 3	13.18	0.01
Número de procedimentos	57.98	0.021
Mudança do tipo de DCEI: entre o Procedimento 3 e Procedimento 4	8.31	0.024
Infarto do miocárdio prévio / Doença arterial coronariana	5.40	0.026
Tipo de Reoperação 4	8.15	0.045
Hipertensão arterial	3.33	0.063
Endocardite prévia	3.43	0.074
Óbito intraoperatório 5	3.23	0.078
Tipo de Dispositivo ao final do procedimento 4	8.31	0.1
Readmissões pós-T0 com diárias de UTI	8.52	0.166
Doença pulmonar obstrutiva crônica	1.69	0.258
Mudança do tipo de DCEI: entre o Procedimento 4 e Procedimento 5	3.21	0.261
Fibrilação / flutter atrial	0.92	0.33
Valvopatias/ Prótese valvares	1.20	0.33
Sexo	0.91	0.337
Tipo de Reoperação 5	3.21	0.385
Óbito intraoperatório 6	1.37	0.395

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

Variable	Statistic	p-value
Tipo de Dispositivo ao final do procedimento 5	3.21	0.434
Mudança do tipo de DCEI: entre o Procedimento 5 e Procedimento 6	1.37	0.481
Óbito intraoperatório 7	0.53	0.687
Tipo de Reoperação 6	1.32	0.692
Parada cardíaca prévia/ Taquicardia ventricular instável	0.22	0.713
Tipo de Dispositivo ao final do procedimento 6	1.37	0.741
Estado de residência	7.01	0.989
Transplante cardíaco prévio	0.13	> 0.999
Acidente Vascular Cerebral/ Acidente isquêmico transitório prévios	0.01	> 0.999
Neoplasia em tratamento ou tratada recentemente	0.01	> 0.999
Tipo de Reoperação 7	0.53	> 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
Readmissões pós-T0 com diálise	0.37	> 0.999