

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_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 em até 30 dias após a alta T0	0.0	< 0.001
Readmissão em até 30 dias	247707.5	< 0.001
Readmissão entre 31 a 60 dias	258535.0	< 0.001
Óbito	87543.0	< 0.001
Óbito durante algum episódio de readmissão hospitalar	288115.5	< 0.001
Readmissão entre 61 a 180 dias	279373.5	< 0.001
Tempo de sobrevivência	117219.5	< 0.001
Tempo de seguimento total	1052783.5	< 0.001
Readmissão em até 1 ano	297940.0	< 0.001
Culturas	422590.5	< 0.001
Equipe Multiprofissional	347927.5	< 0.001
Ultrassom	453236.0	< 0.001
Ecocardiograma	375426.0	< 0.001
Insulina	474610.0	< 0.001
Angiografia	542503.0	< 0.001
Cateter venoso central	500191.5	< 0.001
Citologias	535800.0	< 0.001
Quantidade de exames diagnóstico por imagem	360482.0	< 0.001
Tomografia	464747.5	< 0.001
Ventilação não invasiva	542648.5	< 0.001
Psicofármacos	400549.5	< 0.001
Radiografias	377773.5	< 0.001
Holter	472834.0	< 0.001
Quantidade de exames por métodos gráficos	384035.5	< 0.001
ECC	385531.5	< 0.001
Diuretico	411150.5	< 0.001
Exames laboratoriais	396047.0	< 0.001
Quantidade de exames de análises clínicas	396090.0	< 0.001
DVA	451345.5	< 0.001
Núm. de hospitalizações pré-procedimento	432631.0	< 0.001
Quantidade de procedimentos invasivos	466097.0	< 0.001

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

Variable	Statistic	p-value
Número da Admissão T0	451851.5	< 0.001
Número de comorbidades	421848.5	< 0.001
Vasodilator	454288.0	< 0.001
Interconsulta médica	488422.0	< 0.001
Transusão de hemoderivados	537840.5	< 0.001
Broncodilator	527848.0	< 0.001
Quantidade de exames histopatológicos	538083.5	< 0.001
Intervenção cardiovascular em laboratório de hemodinâmica	543744.5	< 0.001
Idade no momento do primeiro procedimento	435709.0	0.001
Idade no Procedimento 1	435709.0	0.001
Antagonista da Aldosterona	480138.0	0.002
Antiplaquetario EV	543955.0	0.002
Quantidade de classes medicamentosas utilizadas	440657.5	0.002
Bomba de infusão contínua	533417.0	0.002
Antifúngicos	533523.5	0.002
Antiarrítmicos	488637.0	0.002
Antibióticos	449079.5	0.003
Ressonancia magnetica	515870.5	0.005
Cintilografia	526855.0	0.008
Dieta enteral	550345.0	0.008
Antihipertensivo	527829.5	0.012
Ano do procedimento 1	465555.0	0.016
Ano da admissão T0	466054.5	0.018
Número de atendimentos	474029.0	0.021
Tempo entre o P1 e P2	9390.5	0.022
Betabloqueador	516312.5	0.024
Hormonio tireoidiano	540729.5	0.025
Número de Mudanças do tipo de DCEI	2833.5	0.033
Flebografia	541562.5	0.043
Angio TC	538000.5	0.046
Anticoagulantes orais	532628.0	0.06
Cateterismo	524388.5	0.072
Estatinas	504507.0	0.081
Antiviral	551150.5	0.086
Diárias no serviço de Emergência na admissão T0	529789.0	0.087
Antiplaquetario VO	539891.5	0.098
Cirurgia Cardiovascular	541368.0	0.148
PET-CT	551631.5	0.165
Outros procedimentos cirúrgicos	537984.5	0.18
Insuficiência cardíaca	522501.0	0.182
Tempo entre o P2 e P3	1288.5	0.194
Núm. de hospitalizações pós-procedimento	514429.0	0.207
Eletrofisiologia	543688.0	0.239
Idade no Procedimento 3	1209.5	0.266
Ano do procedimento 2	2696.0	0.275
Ano do procedimento 3	299.0	0.304
Cavografia	552766.0	0.367
Anticonvulsivante	548185.5	0.387
Teste de esforço	561929.5	0.443
Intervenção coronária percutânea	561219.5	0.481
Marca-passo temporário	553513.5	0.493
Biopsias	559764.0	0.578
Espirometria / Ergoespirometria	559373.5	0.611
Drenagem de tórax e punção pericárdica ou pleural	559160.5	0.631

Table 2: Mann-Whitney Test (continued)

Variable	Statistic	p-value
Exames endoscópicos	554425.0	0.637
Cardioversão/ Desfibrilação	559018.5	0.644
Suporte cardiocirculatório	558628.0	0.686
Angio RM	558521.5	0.699
Tilt Test	558450.5	0.708
Transplante cardíaco	558379.5	0.717
Cirurgia Toracica	558273.0	0.732
Aortografia	558237.5	0.737
Instalação de CEC	555145.5	0.74
IECA/BRA	546960.0	0.766
Hipoglicemiante	560855.0	0.773
Angioplastia	557953.5	0.782
Polissonografia	557918.0	0.788
Traqueostomia	557811.5	0.809
Trombolitico	557634.0	0.849
Antiretroviral	557634.0	0.849
Arteriografia	557598.5	0.859
Dieta parenteral	557527.5	0.881
Idade no Procedimento 2	5080.5	0.91
Digoxina	557156.0	0.988
Bloqueador do canal de calcio	557248.0	0.99
Óbito hospitalar	557350.0	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
Stent	557350.0	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))
  }
}

## 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
Insuficiência renal crônica	31.53	< 0.001
Óbito intraoperatório 2	26.19	< 0.001
Admissão em até 180 dias antes da T0	18.80	< 0.001
UTI durante a admissão T0	3886.56	< 0.001
Desfecho final do estudo	372.04	< 0.001
Causa do óbito	694.23	< 0.001
Tipo de Dispositivo ao final do procedimento 2	26.36	< 0.001
Escolaridade	26.30	0.001
Hipertensão arterial	8.83	0.001
Tipo de Reoperação 2	26.55	0.001
Mudança do tipo de DCEI: entre o Procedimento 1 e Procedimento 2	26.15	0.001
Diabetes melittus	12.29	0.002
Insuficiência cardíaca	9.74	0.002
UTI durante os episódios de hospitalização	11.54	0.003
Tipo de Procedimento 1	9.38	0.004
Tipo de Dispositivo ao final do procedimento 1	15.60	0.004
Doença cardíaca	14.47	0.004
Valvopatias/ Prótese valvares	8.86	0.007
Infarto do miocárdio prévio / Doença arterial coronariana	7.16	0.012
Tipo de Reoperação 1	10.74	0.019
Ventilação mecânica / IOT	6.21	0.02
Óbito intraoperatório 3	5.30	0.026
Mudança do tipo de DCEI: entre o Procedimento 2 e Procedimento 3	7.52	0.035
Classe funcional de IC	20.23	0.043
Doença cardíaca	19.73	0.045
Fibrilação / flutter atrial	4.33	0.049
Número de procedimentos	26.31	0.07
Sexo	3.23	0.074
Hemodiálise	9.83	0.082
Neoplasia em tratamento ou tratada recentemente	4.36	0.11
Estado de residência	50.00	0.143
Tipo de Reoperação 3	4.92	0.16
Óbito intraoperatório 4	2.26	0.189
Tipo de Dispositivo ao final do procedimento 3	5.37	0.218
Mudança do tipo de DCEI: entre o Procedimento 3 e Procedimento 4	2.26	0.271
Tipo de Reoperação 4	2.22	0.45
Tipo de Dispositivo ao final do procedimento 4	2.26	0.573
Readmissões pós-T0 com diárias de UTI	0.75	0.606
Óbito intraoperatório 5	0.88	0.64
Mudança do tipo de DCEI: entre o Procedimento 4 e Procedimento 5	0.87	0.642
Acidente Vascular Cerebral/ Acidente isquêmico transitório prévios	0.26	0.732
Parada cardíaca prévia/ Taquicardia ventricular instável	0.04	0.868
Raça	0.98	0.97
Transplante cardíaco prévio	0.19	> 0.999
Endocardite prévia	0.27	> 0.999
Doença pulmonar obstrutiva crônica	0.00	> 0.999

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

Variable	Statistic	p-value
Tipo de Reoperação 5	0.87	> 0.999
Tipo de Dispositivo ao final do procedimento 5	0.87	> 0.999
Tipo de Reoperação 6	0.36	> 0.999
Tipo de Dispositivo ao final do procedimento 6	0.37	> 0.999
Óbito intraoperatório 6	0.37	> 0.999
Tipo de Reoperação 7	0.14	> 0.999
Tipo de Dispositivo ao final do procedimento 7	0.14	> 0.999
Óbito intraoperatório 7	0.14	> 0.999
Tipo de Reoperação 8	0.05	> 0.999
Tipo de Dispositivo ao final do procedimento 8	0.05	> 0.999
Óbito intraoperatório 8	0.05	> 0.999
Tipo de Reoperação 9	0.02	> 0.999
Tipo de Dispositivo ao final do procedimento 9	0.02	> 0.999
Óbito intraoperatório 9	0.02	> 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 5 e Procedimento 6	0.37	> 0.999
Mudança do tipo de DCEI: entre o Procedimento 6 e Procedimento 7	0.14	> 0.999
Mudança do tipo de DCEI: entre o Procedimento 7 e Procedimento 8	0.05	> 0.999
Mudança do tipo de DCEI: entre o Procedimento 8 e Procedimento 9	0.02	> 0.999
Mudança do tipo de DCEI: entre o Procedimento 9 e Procedimento 10	0.00	> 0.999
Diálise durante os episódios de hospitalização	0.21	> 0.999
Diálise durante a admissão T0	0.11	> 0.999
Readmissões pós-T0 com diálise	0.10	> 0.999
Óbito intraoperatório 1	NaN	NaN
Desfecho principal da admissão T0	NaN	NaN
Óbito intraoperatório	NaN	NaN