

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
Núm. de hospitalizações pós-procedimento	2309643.0	< 0.001
Readmissão em até 30 dias	2305952.5	< 0.001
Readmissão entre 31 a 60 dias	0.0	< 0.001
Readmissão entre 61 a 180 dias	267672.0	< 0.001
Readmissão em até 1 ano	506160.0	< 0.001
Número de atendimentos	2708238.5	< 0.001
Óbito em até 180 dias após a alta T0	5645265.0	< 0.001
Óbito durante algum episódio de readmissão hospitalar	5389933.0	< 0.001
Tempo entre o P1 e P2	1374007.0	< 0.001
Óbito em até 30 dias após a alta T0	6476889.0	< 0.001
Óbito	5579965.5	< 0.001
Número da Admissão T0	5614153.0	< 0.001
Quantidade de exames diagnóstico por imagem	5338225.0	< 0.001
Antiarrítmicos	6004678.5	< 0.001
Radiografias	5478434.5	< 0.001
Equipe Multiprofissional	5658686.0	< 0.001
Ecocardiograma	5746426.0	< 0.001
Ultrassom	6245838.0	< 0.001
DVA	5947448.0	< 0.001
Antagonista da Aldosterona	6013213.5	< 0.001
Quantidade de exames por métodos gráficos	5571128.0	< 0.001
Transplante cardíaco	6796625.0	< 0.001
ECG	5600033.5	< 0.001
Diuretico	5780235.5	< 0.001
Biopsias	6744263.0	< 0.001
Exames laboratoriais	5670440.0	< 0.001
Quantidade de exames de análises clínicas	5670887.0	< 0.001
Tempo de sobrevida	330780.0	< 0.001
Núm. de hospitalizações pré-procedimento	5908957.5	< 0.001
Tempo de seguimento total	8100270.5	< 0.001
Quantidade de procedimentos invasivos	6153279.0	< 0.001
Ano do procedimento 2	1044185.5	< 0.001
Quantidade de classes medicamentosas utilizadas	5787420.0	< 0.001
Culturas	6272125.0	< 0.001
Insuficiência cardíaca	6130742.5	< 0.001
Suporte cardiocirculatório	6795398.0	< 0.001

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

Variable	Statistic	p-value
Ressonancia magnetica	6486680.5	< 0.001
Óbito em até 1 ano após a alta T0	5353381.5	< 0.001
Quantidade de exames histopatológicos	6716011.5	< 0.001
Cateterismo	6386912.0	< 0.001
Vasodilator	6125383.0	< 0.001
Anticoagulantes orais	6538163.5	< 0.001
Psicofármacos	6081817.5	< 0.001
Exames endoscópicos	6719483.0	< 0.001
Número de comorbidades	5970737.0	< 0.001
Holter	6457291.0	< 0.001
Antiviral	6799930.5	< 0.001
Antifúngicos	6673320.5	< 0.001
Cateter venoso central	6634741.0	< 0.001
Digoxina	6577086.5	< 0.001
Cintilografia	6627743.5	< 0.001
Tomografia	6471278.5	< 0.001
Bloqueador do canal de calcio	6717613.0	< 0.001
Betabloqueador	6497585.0	< 0.001
Diárias no serviço de Emergência na admissão T0	6549617.5	< 0.001
Instalação de CEC	6759027.0	< 0.001
Estatinas	6304458.5	< 0.001
Eletrofisiologia	6686872.0	< 0.001
Broncodilator	6732883.0	< 0.001
Outros procedimentos cirúrgicos	6633470.0	< 0.001
IECA/BRA	6293080.0	< 0.001
Antibióticos	6257379.0	< 0.001
Ano do procedimento 3	124289.5	< 0.001
Bomba de infusão contínua	6771810.0	< 0.001
Tempo entre o P2 e P3	121476.5	< 0.001
Insulina	6673531.5	< 0.001
Anticonvulsivante	6736354.0	< 0.001
Óbito hospitalar	7014648.0	< 0.001
Óbito em até 2 anos após a alta T0	5141537.0	< 0.001
Cardioversão/ Desfibrilação	6845855.0	< 0.001
Transfusão de hemoderivados	6817410.0	< 0.001
Citologias	6838611.5	0.001
Angio TC	6782464.5	0.001
Angio RM	6862594.0	0.003
Idade no Procedimento 2	899411.5	0.004
Idade no momento do primeiro procedimento	7282330.5	0.004
Idade no Procedimento 1	7282330.5	0.004
Espirometria / Ergoespirometria	6857044.0	0.006
Intervenção coronária percutânea	6843718.0	0.01
Teste de esforço	6841747.0	0.011
Tilt Test	6869232.0	0.012
Antiplaquetario EV	6855845.5	0.012
Arteriografia	6882788.0	0.018
Cirurgia Toracica	6870117.5	0.019
PET-CT	6862490.0	0.034
Interconsulta médica	6749579.5	0.038
Angioplastia	6879337.5	0.044
Antihipertensivo	6819910.5	0.072
Dieta enteral	6877703.0	0.096

Table 2: Mann-Whitney Test (continued)

Variable	Statistic	p-value
Ventilação não invasiva	6912960.0	0.125
Flebografia	6855592.0	0.157
Óbito em até 3 anos após a alta T0	4962989.5	0.163
Marca-passo temporário	6867946.0	0.172
Intervenção cardiovascular em laboratório de hemodinâmica	6875375.0	0.175
Ano da admissão T0	7047783.5	0.176
Ano do procedimento 1	7069755.0	0.198
Hormonio tireoidiano	6868473.0	0.313
Polissonografia	6887345.0	0.32
Trombolítico	6900192.0	0.415
Drenagem de tórax e punção pericárdica ou pleural	6905573.0	0.461
Antiretroviral	6891264.5	0.482
Dieta parenteral	6898368.0	0.516
Número de Mudanças do tipo de DCEI	813920.5	0.546
Idade no Procedimento 3	102753.0	0.628
Angiografia	6891000.0	0.659
Aortografia	6891458.0	0.7
Hipoglicemiante	6879391.0	0.712
Antiplaquetario VO	6907659.5	0.739
Stent	6895632.0	0.806
Cirurgia Cardiovascular	6902839.5	0.848
Traqueostomia	6894001.5	0.87
Cavografia	6892677.0	0.893

```

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	27.51	< 0.001
Doença cardíaca	32.28	< 0.001
Infarto do miocárdio prévio / Doença arterial coronariana	16.75	< 0.001
Insuficiência cardíaca	67.16	< 0.001
Fibrilação / flutter atrial	10.16	< 0.001
Transplante cardíaco prévio	48.42	< 0.001
Número de procedimentos	92.12	< 0.001
Tipo de Procedimento 1	76.19	< 0.001
Tipo de Reoperação 1	91.09	< 0.001
Tipo de Dispositivo ao final do procedimento 1	90.46	< 0.001
Tipo de Reoperação 2	558.36	< 0.001
Tipo de Dispositivo ao final do procedimento 2	52.29	< 0.001
Óbito intraoperatório 2	41.96	< 0.001
Tipo de Reoperação 3	67.94	< 0.001
Tipo de Dispositivo ao final do procedimento 3	65.03	< 0.001
Óbito intraoperatório 3	62.72	< 0.001
Tipo de Reoperação 4	54.11	< 0.001
Tipo de Dispositivo ao final do procedimento 4	54.27	< 0.001
Óbito intraoperatório 4	46.93	< 0.001
Mudança do tipo de DCEI: entre o Procedimento 1 e Procedimento 2	41.52	< 0.001
Mudança do tipo de DCEI: entre o Procedimento 2 e Procedimento 3	62.79	< 0.001
Mudança do tipo de DCEI: entre o Procedimento 3 e Procedimento 4	48.36	< 0.001
UTI durante os episódios de hospitalização	143.74	< 0.001
Admissão em até 180 dias antes da T0	104.87	< 0.001
UTI durante a admissão T0	2769.94	< 0.001
Readmissões pós-T0 com diárias de UTI	219.24	< 0.001
Desfecho final do estudo	184.50	< 0.001
Causa do óbito	67.62	< 0.001
Ventilação mecânica / IOT	59.21	< 0.001
Classe funcional de IC	27.93	< 0.001
Desfecho principal da admissão T0	15.81	0.001
Doença cardíaca	15.52	0.004
Díalise durante os episódios de hospitalização	12.23	0.004
Diabetes melittus	6.95	0.007
Parada cardíaca prévia/ Taquicardia ventricular instável	7.23	0.011
Valvopatias/ Prótese valvares	6.51	0.013
Óbito intraoperatório 6	6.50	0.016
Óbito intraoperatório 5	6.41	0.017
Tipo de Reoperação 6	12.39	0.017
Tipo de Reoperação 7	16.24	0.017
Mudança do tipo de DCEI: entre o Procedimento 4 e Procedimento 5	8.28	0.021
Mudança do tipo de DCEI: entre o Procedimento 5 e Procedimento 6	8.63	0.022
Hemodiálise	6.41	0.033
Tipo de Dispositivo ao final do procedimento 5	10.98	0.036
Díalise durante a admissão T0	51.51	0.039
Tipo de Reoperação 5	7.68	0.051
Tipo de Dispositivo ao final do procedimento 6	9.68	0.06
Readmissões pós-T0 com diálise	16.71	0.071
Mudança do tipo de DCEI: entre o Procedimento 6 e Procedimento 7	5.60	0.097
Insuficiência renal crônica	2.85	0.102
Óbito intraoperatório 7	2.77	0.104
Endocardite prévia	2.35	0.137
Estado de residência	37.04	0.143

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

Variable	Statistic	p-value
Tipo de Dispositivo ao final do procedimento 7	5.22	0.177
Acidente Vascular Cerebral/ Acidente isquêmico transitório prévios	0.99	0.333
Neoplasia em tratamento ou tratada recentemente	0.93	0.412
Doença pulmonar obstrutiva crônica	0.56	0.478
Hipertensão arterial	0.49	0.502
Tipo de Dispositivo ao final do procedimento 8	0.72	0.63
Tipo de Reoperação 8	0.72	0.637
Óbito intraoperatório	0.66	0.658
Óbito intraoperatório 8	0.72	0.658
Mudança do tipo de DCEI: entre o Procedimento 7 e Procedimento 8	0.72	0.672
Raça	2.06	0.88
Sexo	0.01	0.927
Óbito intraoperatório 1	0.42	> 0.999
Tipo de Reoperação 9	0.30	> 0.999
Tipo de Dispositivo ao final do procedimento 9	0.30	> 0.999
Óbito intraoperatório 9	0.30	> 0.999
Tipo de Reoperação 10	0.06	> 0.999
Tipo de Dispositivo ao final do procedimento 10	0.06	> 0.999
Óbito intraoperatório 10	0.06	> 0.999
Mudança do tipo de DCEI: entre o Procedimento 8 e Procedimento 9	0.30	> 0.999
Mudança do tipo de DCEI: entre o Procedimento 9 e Procedimento 10	0.06	> 0.999