

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_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 1 ano após a alta T0 | 0.0 | < 0.001 |
| Óbito | 188870.0 | < 0.001 |
| Óbito durante algum episódio de readmissão hospitalar | 610546.5 | < 0.001 |
| Readmissão em até 1 ano | 568564.0 | < 0.001 |
| Tempo de sobrevida | 228797.0 | < 0.001 |
| Tempo de seguimento total | 2434948.5 | < 0.001 |
| Readmissão entre 61 a 180 dias | 944291.0 | < 0.001 |
| Insuficiência cardíaca | 1081777.5 | < 0.001 |
| Diuretico | 1012226.5 | < 0.001 |
| Núm. de hospitalizações pós-procedimento | 997134.0 | < 0.001 |
| Readmissão entre 31 a 60 dias | 1249598.5 | < 0.001 |
| DVA | 1117039.5 | < 0.001 |
| Número de atendimentos | 1003631.5 | < 0.001 |
| Antagonista da Aldosterona | 1148327.0 | < 0.001 |
| Ultrassom | 1241310.0 | < 0.001 |
| Número de comorbidades | 1013617.0 | < 0.001 |
| Quantidade de classes medicamentosas utilizadas | 1010465.5 | < 0.001 |
| Quantidade de exames diagnóstico por imagem | 1014285.5 | < 0.001 |
| Exames laboratoriais | 1033253.5 | < 0.001 |
| Quantidade de exames de análises clínicas | 1033762.5 | < 0.001 |
| ECG | 1025191.0 | < 0.001 |
| Quantidade de exames por métodos gráficos | 1027830.5 | < 0.001 |
| Equipe Multiprofissional | 1090169.5 | < 0.001 |
| Radiografias | 1044659.0 | < 0.001 |
| Número da Admissão T0 | 1138599.0 | < 0.001 |
| Digoxina | 1299073.5 | < 0.001 |
| Antiarrítmicos | 1217653.0 | < 0.001 |
| Readmissão em até 30 dias | 1318121.5 | < 0.001 |
| Culturas | 1252307.0 | < 0.001 |
| Tomografia | 1280861.5 | < 0.001 |
| Núm. de hospitalizações pré-procedimento | 1180286.5 | < 0.001 |
| Antibióticos | 1138739.5 | < 0.001 |
| Tempo entre o P1 e P2 | 48915.5 | < 0.001 |
| Anticoagulantes orais | 1328217.0 | < 0.001 |

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

| Variable | Statistic | p-value |
|---|-----------|---------|
| Psicofármacos | 1196676.5 | < 0.001 |
| Citologias | 1399935.0 | < 0.001 |
| Ecocardiograma | 1223636.5 | < 0.001 |
| Bomba de infusão contínua | 1375797.0 | < 0.001 |
| Cintilografia | 1352798.0 | < 0.001 |
| IECA/BRA | 1219693.5 | < 0.001 |
| Holter | 1334929.0 | < 0.001 |
| Quantidade de procedimentos invasivos | 1306682.5 | 0.001 |
| Cirurgia Toracica | 1413139.5 | 0.002 |
| Vasodilator | 1291517.0 | 0.002 |
| Ressonancia magnetica | 1359698.0 | 0.004 |
| PET-CT | 1408608.5 | 0.006 |
| Exames endoscópicos | 1400202.0 | 0.007 |
| Angio TC | 1387997.0 | 0.012 |
| Tempo entre o P2 e P3 | 4955.0 | 0.016 |
| Quantidade de exames histopatológicos | 1405711.5 | 0.02 |
| Estatinas | 1319044.0 | 0.025 |
| Antifúngicos | 1399706.5 | 0.029 |
| Antiplaquetario VO | 1393210.0 | 0.046 |
| Antiviral | 1415025.5 | 0.047 |
| Idade no momento do primeiro procedimento | 1315760.5 | 0.068 |
| Idade no Procedimento 1 | 1315760.5 | 0.068 |
| Ano do procedimento 1 | 1316252.5 | 0.069 |
| Ano da admissão T0 | 1310455.5 | 0.077 |
| Cateterismo | 1376925.0 | 0.087 |
| Cateter venoso central | 1399511.0 | 0.098 |
| Outros procedimentos cirúrgicos | 1388690.0 | 0.098 |
| Interconsulta médica | 1476607.5 | 0.102 |
| Espirometria / Ergoespirometria | 1416046.0 | 0.106 |
| Ano do procedimento 2 | 33490.0 | 0.131 |
| Flebografia | 1407822.0 | 0.133 |
| Dieta enteral | 1420678.0 | 0.169 |
| Ano do procedimento 3 | 4083.0 | 0.17 |
| Angiografia | 1420773.0 | 0.185 |
| Teste de esforço | 1438310.5 | 0.207 |
| Ventilação não invasiva | 1421055.5 | 0.236 |
| Transfusão de hemoderivados | 1436440.5 | 0.246 |
| Tilt Test | 1421240.5 | 0.27 |
| Insulina | 1402983.0 | 0.299 |
| Idade no Procedimento 2 | 31045.0 | 0.328 |
| Angio RM | 1421606.0 | 0.338 |
| Biopsias | 1432420.0 | 0.367 |
| Antihipertensivo | 1410280.0 | 0.386 |
| Intervenção cardiovascular em laboratório de hemodinâmica | 1431765.5 | 0.394 |
| Antiplaquetario EV | 1431672.0 | 0.398 |
| Suporte cardiocirculatório | 1429334.5 | 0.524 |
| Cardioversão/ Desfibrilação | 1422838.5 | 0.551 |
| Transplante cardíaco | 1428867.0 | 0.558 |
| Drenagem de tórax e punção pericárdica ou pleural | 1423093.5 | 0.593 |
| Betabloqueador | 1441448.5 | 0.596 |
| Aortografia | 1428306.0 | 0.603 |
| Angioplastia | 1427745.0 | 0.658 |
| Polissonografia | 1427651.5 | 0.668 |

Table 2: Mann-Whitney Test (continued)

| Variable | Statistic | p-value |
|---|-----------|---------|
| Traqueostomia | 1427464.5 | 0.69 |
| Eletrofisiologia | 1432887.5 | 0.717 |
| Trombolítico | 1426997.0 | 0.754 |
| Antiretroviral | 1426997.0 | 0.754 |
| Cirurgia Cardiovascular | 1420800.0 | 0.755 |
| Marca-passo temporário | 1428792.5 | 0.773 |
| Arteriografia | 1426810.0 | 0.786 |
| Idade no Procedimento 3 | 3147.0 | 0.792 |
| Dieta parenteral | 1426716.5 | 0.805 |
| Bloqueador do canal de calcio | 1423246.5 | 0.808 |
| Intervenção coronária percutânea | 1428076.5 | 0.83 |
| Hormonio tireoidiano | 1428650.5 | 0.836 |
| Hipoglicemiante | 1422652.0 | 0.851 |
| Broncodilator | 1428235.0 | 0.885 |
| Diárias no serviço de Emergência na admissão T0 | 1423725.0 | 0.92 |
| Número de Mudanças do tipo de DCEI | 26287.0 | 0.921 |
| Anticonvulsivante | 1424895.5 | 0.935 |
| Cavografia | 1426850.0 | 0.94 |
| Instalação de CEC | 1425489.5 | 0.943 |
| Óbito hospitalar | 1426249.0 | NaN |
| Óbito em até 30 dias após a alta T0 | 1426249.0 | NaN |
| Óbito em até 180 dias após a alta T0 | 1426249.0 | NaN |
| Óbito em até 2 anos após a alta T0 | NaN | NaN |
| Óbito em até 3 anos após a alta T0 | NaN | NaN |
| Stent | 1426249.0 | 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 cardíaca | 49.94 | < 0.001 |
| Valvopatias/ Prótese valvares | 42.48 | < 0.001 |
| Insuficiência renal crônica | 21.55 | < 0.001 |
| Tipo de Dispositivo ao final do procedimento 1 | 41.28 | < 0.001 |
| Tipo de Reoperação 2 | 61.57 | < 0.001 |
| Tipo de Dispositivo ao final do procedimento 2 | 59.23 | < 0.001 |
| Óbito intraoperatório 2 | 56.85 | < 0.001 |
| Mudança do tipo de DCEI: entre o Procedimento 1 e Procedimento 2 | 56.84 | < 0.001 |
| UTI durante a admissão T0 | 2400.52 | < 0.001 |
| Desfecho final do estudo | 1135.07 | < 0.001 |
| Causa do óbito | 1402.52 | < 0.001 |
| Ventilação mecânica / IOT | 45.86 | < 0.001 |
| Óbito intraoperatório 3 | 11.92 | < 0.001 |
| Classe funcional de IC | 32.81 | 0.003 |
| Díálise durante os episódios de hospitalização | 25.89 | 0.003 |
| UTI durante os episódios de hospitalização | 9.09 | 0.003 |
| Tipo de Procedimento 1 | 6.58 | 0.006 |
| Tipo de Dispositivo ao final do procedimento 3 | 16.06 | 0.006 |
| Mudança do tipo de DCEI: entre o Procedimento 2 e Procedimento 3 | 12.44 | 0.007 |
| Doença cardíaca | 27.89 | 0.007 |
| Diabetes melittus | 6.72 | 0.008 |
| Sexo | 7.56 | 0.008 |
| Doença cardíaca | 12.27 | 0.008 |
| Tipo de Reoperação 1 | 14.73 | 0.012 |
| Tipo de Reoperação 3 | 12.04 | 0.013 |
| Fibrilação / flutter atrial | 6.27 | 0.014 |
| Doença pulmonar obstrutiva crônica | 8.61 | 0.014 |
| Número de procedimentos | 57.36 | 0.014 |
| Óbito intraoperatório 4 | 6.12 | 0.016 |
| Díálise durante a admissão T0 | 96.51 | 0.023 |
| Hipertensão arterial | 5.02 | 0.027 |
| Infarto do miocárdio prévio / Doença arterial coronariana | 4.54 | 0.035 |
| Admissão em até 180 dias antes da T0 | 3.95 | 0.049 |
| Readmissões pós-T0 com diálise | 15.88 | 0.058 |
| Mudança do tipo de DCEI: entre o Procedimento 3 e Procedimento 4 | 6.12 | 0.062 |
| Parada cardíaca prévia/ Taquicardia ventricular instável | 3.04 | 0.09 |
| Tipo de Reoperação 4 | 6.01 | 0.117 |
| Escolaridade | 9.88 | 0.121 |
| Hemodiálise | 4.12 | 0.152 |
| Óbito intraoperatório 5 | 2.38 | 0.173 |
| Tipo de Dispositivo ao final do procedimento 4 | 6.12 | 0.185 |
| Mudança do tipo de DCEI: entre o Procedimento 4 e Procedimento 5 | 2.37 | 0.257 |
| Tipo de Reoperação 5 | 2.37 | 0.431 |
| Óbito intraoperatório 6 | 1.01 | 0.623 |
| Tipo de Dispositivo ao final do procedimento 5 | 2.37 | 0.626 |
| Mudança do tipo de DCEI: entre o Procedimento 5 e Procedimento 6 | 1.01 | 0.644 |
| Endocardite prévia | 0.22 | 0.724 |
| Raça | 2.86 | 0.738 |

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

| Variable | Statistic | p-value |
|--|-----------|---------|
| Estado de residência | 13.25 | 0.765 |
| Tipo de Reoperação 6 | 0.97 | 0.771 |
| Acidente Vascular Cerebral/ Acidente isquêmico transitório prévios | 0.10 | 0.83 |
| Transplante cardíaco prévio | 0.50 | > 0.999 |
| Neoplasia em tratamento ou tratada recentemente | 0.07 | > 0.999 |
| Tipo de Dispositivo ao final do procedimento 6 | 1.01 | > 0.999 |
| Tipo de Reoperação 7 | 0.39 | > 0.999 |
| Tipo de Dispositivo ao final do procedimento 7 | 0.38 | > 0.999 |
| Óbito intraoperatório 7 | 0.39 | > 0.999 |
| Tipo de Reoperação 8 | 0.15 | > 0.999 |
| Tipo de Dispositivo ao final do procedimento 8 | 0.15 | > 0.999 |
| Óbito intraoperatório 8 | 0.15 | > 0.999 |
| Tipo de Reoperação 9 | 0.06 | > 0.999 |
| Tipo de Dispositivo ao final do procedimento 9 | 0.06 | > 0.999 |
| Óbito intraoperatório 9 | 0.06 | > 0.999 |
| Tipo de Reoperação 10 | 0.01 | > 0.999 |
| Tipo de Dispositivo ao final do procedimento 10 | 0.01 | > 0.999 |
| Óbito intraoperatório 10 | 0.01 | > 0.999 |
| Mudança do tipo de DCEI: entre o Procedimento 6 e Procedimento 7 | 0.38 | > 0.999 |
| Mudança do tipo de DCEI: entre o Procedimento 7 e Procedimento 8 | 0.15 | > 0.999 |
| Mudança do tipo de DCEI: entre o Procedimento 8 e Procedimento 9 | 0.06 | > 0.999 |
| Mudança do tipo de DCEI: entre o Procedimento 9 e Procedimento 10 | 0.01 | > 0.999 |
| Readmissões pós-T0 com diárias de UTI | 0.54 | > 0.999 |
| Óbito intraoperatório 1 | NaN | NaN |
| Desfecho principal da admissão T0 | NaN | NaN |
| Óbito intraoperatório | NaN | NaN |