

Correlations

Eduardo Yuki Yada

Imports

```
library(tidyverse)
library(yaml)
library(kableExtra)
library(ggcorrplot)
```

Loading data

```
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(.): the standard deviation is 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 | 0.85 |
| Idade no momento do primeiro procedimento | Idade no Procedimento 2 | 0.85 |
| Ano do procedimento 1 | Ano da admissão T0 | 0.85 |
| Idade no Procedimento 1 | Idade no momento do primeiro procedimento | 0.85 |
| Idade no Procedimento 1 | Idade no Procedimento 2 | 0.85 |
| Idade no Procedimento 2 | Idade no momento do primeiro procedimento | 0.85 |
| Idade no Procedimento 2 | Idade no Procedimento 1 | 0.85 |
| Número de atendimentos | Núm. de hospitalizações pós-procedimento | 0.85 |
| Número da Admissão T0 | Núm. de hospitalizações pré-procedimento | 0.85 |
| Núm. de hospitalizações pós-procedimento | Número de atendimentos | 0.85 |
| Núm. de hospitalizações pré-procedimento | Número da Admissão T0 | 0.85 |
| Ano da admissão T0 | Ano do procedimento 1 | 0.85 |
| Readmissão em até 30 dias | Readmissão entre 31 a 60 dias | 0.85 |
| Readmissão entre 31 a 60 dias | Readmissão em até 30 dias | 0.85 |
| Readmissão entre 61 a 180 dias | Readmissão em até 1 ano | 0.85 |
| Readmissão em até 1 ano | Readmissão entre 61 a 180 dias | 0.85 |
| Óbito durante algum episódio de readmissão hospitalar | Óbito | 0.85 |
| Óbito | Óbito durante algum episódio de readmissão hospitalar | 0.85 |
| DVA | Quantidade de medicamentos de ação cardiovascular | 0.85 |
| Diuretico | Quantidade de medicamentos de ação cardiovascular | 0.85 |
| Vasodilator | Quantidade de medicamentos de ação cardiovascular | 0.85 |
| Antibióticos | Quantidade de antimicrobianos | 0.85 |
| Quantidade de classes medicamentosas utilizadas | Quantidade de classes medicamentosas de ação cardiovascular | 0.85 |
| Quantidade de classes medicamentosas de ação cardiovascular | Quantidade de classes medicamentosas utilizadas | 0.85 |

Table 1: Pearson Correlation (*continued*)

| row | column | cor |
|---|---|-----|
| Quantidade de medicamentos de ação cardiovascular | DVA | |
| Quantidade de medicamentos de ação cardiovascular | Diuretico | |
| Quantidade de medicamentos de ação cardiovascular | Vasodilator | |
| Quantidade de antimicrobianos | Antibióticos | |
| Stent | Angioplastia | |
| Stent | Interconsulta médica | |
| Angioplastia | Stent | |
| Interconsulta médica | Stent | |
| Equipe Multiprofissional | Radiografias | |
| Equipe Multiprofissional | Quantidade de exames diagnóstico por imagem | |
| ECG | Quantidade de exames por métodos gráficos | |
| ECG | Exames laboratoriais | |
| ECG | Quantidade de exames de análises clínicas | |
| ECG | Radiografias | |
| ECG | Quantidade de exames diagnóstico por imagem | |
| Quantidade de exames por métodos gráficos | ECG | |
| Quantidade de exames por métodos gráficos | Exames laboratoriais | |
| Quantidade de exames por métodos gráficos | Quantidade de exames de análises clínicas | |
| Quantidade de exames por métodos gráficos | Radiografias | |
| Quantidade de exames por métodos gráficos | Quantidade de exames diagnóstico por imagem | |
| Exames laboratoriais | ECG | |
| Exames laboratoriais | Quantidade de exames por métodos gráficos | |
| Exames laboratoriais | Quantidade de exames de análises clínicas | |
| Exames laboratoriais | Radiografias | |
| Exames laboratoriais | Quantidade de exames diagnóstico por imagem | |
| Quantidade de exames de análises clínicas | ECG | |
| Quantidade de exames de análises clínicas | Quantidade de exames por métodos gráficos | |
| Quantidade de exames de análises clínicas | Exames laboratoriais | |
| Quantidade de exames de análises clínicas | Radiografias | |
| Quantidade de exames de análises clínicas | Quantidade de exames diagnóstico por imagem | |
| Biopsias | Quantidade de exames histopatológicos | |
| Quantidade de exames histopatológicos | Biopsias | |
| Radiografias | Equipe Multiprofissional | |
| Radiografias | ECG | |
| Radiografias | Quantidade de exames por métodos gráficos | |
| Radiografias | Exames laboratoriais | |
| Radiografias | Quantidade de exames de análises clínicas | |
| Radiografias | Quantidade de exames diagnóstico por imagem | |
| Quantidade de exames diagnóstico por imagem | Equipe Multiprofissional | |
| Quantidade de exames diagnóstico por imagem | ECG | |
| Quantidade de exames diagnóstico por imagem | Quantidade de exames por métodos gráficos | |
| Quantidade de exames diagnóstico por imagem | Exames laboratoriais | |
| Quantidade de exames diagnóstico por imagem | Quantidade de exames de análises clínicas | |
| Quantidade de exames diagnóstico por imagem | Radiografias | |

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 |
|---|-----------|---------|
| Óbito hospitalar | 0.0 | < 0.001 |
| Óbito | 332216.0 | < 0.001 |
| Antifúngicos | 899398.5 | < 0.001 |
| Culturas | 563509.0 | < 0.001 |
| DVA | 342023.5 | < 0.001 |
| Óbito intraoperatório | 1978131.0 | < 0.001 |
| Tempo de seguimento total | 3877935.5 | < 0.001 |
| Tempo de sobrevida | 438197.5 | < 0.001 |
| Ultrassom | 842750.5 | < 0.001 |
| Ecocardiograma | 437503.5 | < 0.001 |
| Quantidade de exames diagnóstico por imagem | 274117.0 | < 0.001 |
| Citologias | 1316135.0 | < 0.001 |
| Equipe Multiprofissional | 396670.5 | < 0.001 |
| Exames endoscópicos | 1253291.0 | < 0.001 |
| Tomografia | 884614.0 | < 0.001 |
| Radiografias | 362633.0 | < 0.001 |
| Quantidade de exames histopatológicos | 1290229.0 | < 0.001 |
| Diuretico | 478386.5 | < 0.001 |
| Quantidade de exames de análises clínicas | 444687.5 | < 0.001 |
| Exames laboratoriais | 444852.0 | < 0.001 |
| Quantidade de exames por métodos gráficos | 476424.0 | < 0.001 |
| ECG | 478209.5 | < 0.001 |
| Quantidade de medicamentos de ação cardiovascular | 455194.5 | < 0.001 |

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

| Variable | Statistic | p-value |
|---|-----------|---------|
| Anticonvulsivante | 1087972.5 | < 0.001 |
| Quantidade de antimicrobianos | 532260.0 | < 0.001 |
| Antibióticos | 535397.0 | < 0.001 |
| Outros procedimentos cirúrgicos | 1095619.0 | < 0.001 |
| Transfusão de hemoderivados | 1321993.5 | < 0.001 |
| Quantidade de classes medicamentosas utilizadas | 385566.5 | < 0.001 |
| Quantidade de procedimentos invasivos | 924613.0 | < 0.001 |
| Psicofármacos | 722847.5 | < 0.001 |
| Antiarrítmicos | 914624.0 | < 0.001 |
| Insulina | 1073381.0 | < 0.001 |
| Núm. de hospitalizações pós-procedimento | 2897827.0 | < 0.001 |
| Quantidade de classes medicamentosas de ação cardiovascular | 453480.5 | < 0.001 |
| Vasodilator | 886315.0 | < 0.001 |
| Ventilação não invasiva | 1430121.5 | < 0.001 |
| Suporte cardiocirculatório | 1430168.0 | < 0.001 |
| Cirurgia Torácica | 1442774.5 | < 0.001 |
| Dieta enteral | 1313804.0 | < 0.001 |
| Número de procedimentos na admissão T0 | 1934429.5 | < 0.001 |
| Cateter venoso central | 1356105.0 | < 0.001 |
| Bomba de infusão contínua | 1248648.0 | < 0.001 |
| Instalação de CEC | 1398334.5 | < 0.001 |
| Stent | 1478905.0 | < 0.001 |
| Antiviral | 1331166.5 | < 0.001 |
| Cateterismo | 1260996.5 | < 0.001 |
| Antagonista da Aldosterona | 1095200.0 | < 0.001 |
| Número de comorbidades | 1566295.5 | < 0.001 |
| Cirurgia Cardiovascular | 1359671.5 | < 0.001 |
| Tempo entre o P1 e P2 | 102961.5 | < 0.001 |
| PET-CT | 1440122.5 | < 0.001 |
| Trombolítico | 1357942.5 | < 0.001 |
| Drenagem de tórax e punção pericárdica ou pleural | 1445359.5 | < 0.001 |
| Readmissão em até 1 ano | 2328394.0 | < 0.001 |
| Número da Admissão T0 | 1731511.5 | < 0.001 |
| Insuficiência cardíaca | 1134592.5 | < 0.001 |
| Angio TC | 1393464.0 | < 0.001 |
| Interconsulta médica | 1312191.0 | < 0.001 |
| Dieta parenteral | 1342111.0 | < 0.001 |
| Núm. de hospitalizações pré-procedimento | 1746026.5 | < 0.001 |
| Readmissão entre 61 a 180 dias | 2259881.0 | < 0.001 |
| Holter | 1346142.5 | < 0.001 |
| Biopsias | 1453291.0 | < 0.001 |
| Número de atendimentos | 2384481.5 | < 0.001 |
| Antiplaquetário EV | 1345916.0 | < 0.001 |
| Cavografia | 1450417.0 | < 0.001 |
| Readmissão entre 31 a 60 dias | 2183508.0 | < 0.001 |
| Diárias no serviço de Emergência na admissão T0 | 800251.0 | < 0.001 |
| Cintilografia | 1410912.5 | < 0.001 |
| Óbito durante algum episódio de readmissão hospitalar | 2203598.0 | < 0.001 |
| Intervenção coronária percutânea | 1452095.5 | < 0.001 |
| Traqueostomia | 1473860.5 | < 0.001 |
| Readmissão em até 30 dias | 2143815.0 | 0.001 |
| Polissonografia | 1474217.0 | 0.002 |
| Hipoglicemiante | 1430329.0 | 0.004 |

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

| Variable | Statistic | p-value |
|---|-----------|---------|
| Antihipertensivo | 1325361.5 | 0.007 |
| Anticoagulantes orais | 1323145.0 | 0.011 |
| Cardioversão/ Desfibrilação | 1340416.0 | 0.015 |
| Arteriografia | 1479703.5 | 0.019 |
| Aortografia | 1475129.0 | 0.019 |
| Antiretroviral | 1370231.0 | 0.035 |
| Bloqueador do canal de calcio | 1350046.5 | 0.037 |
| Flebografia | 1458481.5 | 0.038 |
| Tempo entre o P2 e P3 | 7485.5 | 0.039 |
| Ano do procedimento 2 | 70616.5 | 0.073 |
| Digoxina | 1338198.5 | 0.074 |
| Número de procedimentos em até 180 dias | 2086830.0 | 0.093 |
| Betabloqueador | 1328882.0 | 0.102 |
| Número de procedimentos em até 1 ano | 2083948.0 | 0.118 |
| IECA/BRA | 1454860.0 | 0.122 |
| Teste de esforço | 1500126.0 | 0.131 |
| Número de procedimentos em até 60 dias | 2081983.0 | 0.138 |
| Idade no Procedimento 2 | 68187.0 | 0.151 |
| Número de procedimentos em até 30 dias | 2080149.0 | 0.162 |
| Angioplastia | 1480835.0 | 0.21 |
| Transplante cardíaco | 1481299.0 | 0.308 |
| Espirometria / Ergoespirometria | 1491918.0 | 0.317 |
| Eletrofisiologia | 1469523.5 | 0.403 |
| Número de Mudanças do tipo de DCEI | 53175.5 | 0.408 |
| Ano do procedimento 3 | 6060.5 | 0.422 |
| Tilt Test | 1488954.0 | 0.461 |
| Marca-passo temporário | 1347743.0 | 0.494 |
| Angiografia | 1488384.0 | 0.5 |
| Estatinas | 1345525.0 | 0.514 |
| Idade no momento do primeiro procedimento | 2111081.0 | 0.533 |
| Idade no Procedimento 1 | 2111081.0 | 0.533 |
| Ressonancia magnetica | 1500221.0 | 0.537 |
| Ano do procedimento 1 | 2020599.0 | 0.551 |
| Angio RM | 1482674.5 | 0.584 |
| Ano da admissão T0 | 2022566.5 | 0.593 |
| Idade no Procedimento 3 | 4965.5 | 0.867 |
| Intervenção cardiovascular em laboratório de hemodinâmica | 1485872.5 | 0.947 |
| Antiplaquetario VO | 1375472.0 | NaN |
| Hormonio tireoidiano | 1375472.0 | NaN |
| Broncodiltador | 1375472.0 | 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))
}
}

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 | 185.78 | < 0.001 |
| Escolaridade | 100.12 | < 0.001 |
| Doença cardíaca | 84.44 | < 0.001 |
| Classe funcional de IC | 55.90 | < 0.001 |
| Insuficiência cardíaca | 118.04 | < 0.001 |
| Diabetes mellitus | 50.34 | < 0.001 |
| Insuficiência renal crônica | 41.38 | < 0.001 |
| Tipo de Procedimento 1 | 26.46 | < 0.001 |
| Tipo de Reoperação 1 | 51.58 | < 0.001 |
| Tipo de Dispositivo ao final do procedimento 1 | 37.20 | < 0.001 |
| Óbito intraoperatório 1 | 421.31 | < 0.001 |
| Tipo de Reoperação 2 | 68.05 | < 0.001 |
| Tipo de Dispositivo ao final do procedimento 2 | 60.13 | < 0.001 |
| Óbito intraoperatório 2 | 57.05 | < 0.001 |
| Óbito intraoperatório 3 | 260.72 | < 0.001 |
| Mudança do tipo de DCEI: entre o Procedimento 1 e Procedimento 2 | 59.59 | < 0.001 |
| Díálise durante os episódios de hospitalização | 262.02 | < 0.001 |
| UTI durante os episódios de hospitalização | 437.49 | < 0.001 |
| Admissão em até 180 dias antes da T0 | 42.95 | < 0.001 |
| UTI durante a admissão T0 | 9335.69 | < 0.001 |
| Díálise durante a admissão T0 | 716.37 | < 0.001 |
| Desfecho principal da admissão T0 | 15775.35 | < 0.001 |
| Desfecho final do estudo | 1259.05 | < 0.001 |
| Causa do óbito | 1898.29 | < 0.001 |
| Ventilação mecânica / IOT | 194.74 | < 0.001 |
| Mudança do tipo de DCEI: entre o Procedimento 2 e Procedimento 3 | 13.81 | 0.003 |
| Tipo de Reoperação 3 | 13.24 | 0.005 |
| Tipo de Dispositivo ao final do procedimento 3 | 16.21 | 0.006 |
| Hemodiálise | 19.73 | 0.007 |
| Óbito intraoperatório 4 | 8.31 | 0.008 |
| Doença cardíaca | 11.28 | 0.013 |
| Número de procedimentos | 58.08 | 0.02 |
| Infarto do miocárdio prévio / Doença arterial coronariana | 5.44 | 0.023 |
| Mudança do tipo de DCEI: entre o Procedimento 3 e Procedimento 4 | 8.31 | 0.027 |
| Tipo de Reoperação 4 | 8.15 | 0.046 |
| Óbito intraoperatório 5 | 3.23 | 0.07 |
| Hipertensão arterial | 3.30 | 0.075 |
| Endocardite prévia | 3.42 | 0.075 |

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

| Variable | Statistic | p-value |
|--|-----------|---------|
| Tipo de Dispositivo ao final do procedimento 4 | 8.31 | 0.093 |
| Readmissões pós-T0 com diárias de UTI | 8.51 | 0.159 |
| Readmissão cirúrgica em até 30 dias | 2.30 | 0.172 |
| Readmissão cirúrgica em até 1 ano | 2.21 | 0.189 |
| Transplante cardíaco prévio | 2.97 | 0.19 |
| Readmissão cirúrgica entre 61 a 180 dias | 2.50 | 0.193 |
| Mudança do tipo de DCEI: entre o Procedimento 4 e Procedimento 5 | 3.21 | 0.261 |
| Readmissão cirúrgica entre 31 a 60 dias | 1.55 | 0.27 |
| Doença pulmonar obstrutiva crônica | 1.69 | 0.277 |
| Valvopatias/ Prótese valvares | 1.20 | 0.315 |
| Fibrilação / flutter atrial | 0.92 | 0.326 |
| Sexo | 0.92 | 0.357 |
| Tipo de Reoperação 5 | 3.21 | 0.368 |
| Óbito intraoperatório 6 | 1.35 | 0.411 |
| Tipo de Dispositivo ao final do procedimento 5 | 3.21 | 0.445 |
| Mudança do tipo de DCEI: entre o Procedimento 5 e Procedimento 6 | 1.35 | 0.487 |
| Óbito intraoperatório 7 | 0.53 | 0.675 |
| Tipo de Reoperação 6 | 1.30 | 0.693 |
| Parada cardíaca prévia/ Taquicardia ventricular instável | 0.21 | 0.713 |
| Tipo de Dispositivo ao final do procedimento 6 | 1.35 | 0.745 |
| Estado de residência | 7.01 | 0.985 |
| 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 |