

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, font_size = NULL){
  df %>%
    kbl(booktabs = T, longtable = T, caption = caption, digits = digits, format = "latex") %>%
    kable_styling(font_size = font_size,
                  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

pre_columns = df_names %>%
  filter(momento.aquisicao == 'Admissão t0') %>%
  .$variable.name

weird_columns <- c('dieta_parentral', 'dieta_enteral')

eligible_columns <- intersect(na_eligible_columns,
                             unique_eligible_columns) %>%
  intersect(pre_columns)
```

```

eligible_columns <- setdiff(eligible_columns, weird_columns)

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 > 0.9) %>%
  rename_column('row') %>%
  rename_column('column') %>%
  select(row, column, correlation) %>%
  niceFormatting(caption = "Pearson Correlation", font_size = 9)

```

Table 1: Pearson Correlation

row	column	correlation
Idade no momento do primeiro procedimento	Idade no Procedimento 1	1.00
Núm. de hospitalizações pré-procedimento	Número da Admissão T0	0.98
Ano da admissão T0	Ano do procedimento 1	1.00
Antibióticos	Quantidade de antimicrobianos	1.00
Quantidade de classes medicamentosas de ação cardiovascular	Quantidade de classes medicamentosas utilizadas	0.91
Quantidade de procedimentos invasivos	Suporte cardiocirculatório	0.97
ECG	Quantidade de exames por métodos gráficos	1.00
Exames laboratoriais	Radiografias	0.90
Quantidade de exames de análises clínicas	Exames laboratoriais	1.00
Quantidade de exames de análises clínicas	Radiografias	0.90
Quantidade de exames de análises clínicas	Quantidade de exames diagnóstico por imagem	0.93
Quantidade de exames diagnóstico por imagem	Exames laboratoriais	0.93
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	2295262.0	< 0.001
Readmissão em até 30 dias	2290377.0	< 0.001
Readmissão entre 31 a 60 dias	0.0	< 0.001
Readmissão entre 61 a 180 dias	264099.0	< 0.001
Readmissão em até 1 ano	501018.0	< 0.001
Número de procedimentos em até 30 dias	5996083.0	< 0.001
Número de procedimentos em até 60 dias	5882247.5	< 0.001
Número de atendimentos	2704481.5	< 0.001
Óbito durante algum episódio de readmissão hospitalar	5356046.0	< 0.001
Tempo entre o P1 e P2	1374293.0	< 0.001
Quantidade de classes medicamentosas utilizadas	2206162.0	< 0.001
Óbito	5656264.0	< 0.001
Quantidade de exames diagnóstico por imagem	3509134.5	< 0.001
Número da Admissão T0	5594499.0	< 0.001
Radiografias	3634841.0	< 0.001
Quantidade de classes medicamentosas de ação cardiovascular	1892069.0	< 0.001
Quantidade de medicamentos de ação cardiovascular	3207001.0	< 0.001
Quantidade de exames por métodos gráficos	3696397.0	< 0.001
ECG	3719971.5	< 0.001
Equipe Multiprofissional	3823620.0	< 0.001
UTI durante a admissão T0	5852811.0	< 0.001
Antiarrítmicos	3614201.5	< 0.001
Ecocardiograma	3917659.0	< 0.001
Ultrassom	4302707.0	< 0.001
DVA	3570017.5	< 0.001

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

Variable	Statistic	p-value
Exames laboratoriais	3809629.0	< 0.001
Quantidade de exames de análises clínicas	3809813.5	< 0.001
Diuretico	3431767.0	< 0.001
Antagonista da Aldosterona	3621286.0	< 0.001
Tempo de sobrevida	329150.0	< 0.001
Tempo de seguimento total	8049517.5	< 0.001
Núm. de hospitalizações pré-procedimento	5888540.5	< 0.001
Número de procedimentos em até 180 dias	6643962.5	< 0.001
Ano do procedimento 2	1044544.0	< 0.001
Quantidade de procedimentos invasivos	4239531.0	< 0.001
Biopsias	4701221.5	< 0.001
Transplante cardíaco	4744189.5	< 0.001
Insuficiência cardíaca	3717639.5	< 0.001
Culturas	4325300.5	< 0.001
Ressonancia magnetica	4475737.5	< 0.001
Anticoagulantes orais	3998260.0	< 0.001
Vasodilator	3714341.0	< 0.001
Cateterismo	4412941.5	< 0.001
Psicofármacos	3658792.0	< 0.001
Número de comorbidades	5958286.5	< 0.001
Antifúngicos	4097729.0	< 0.001
Suporte cardiocirculatório	4737425.5	< 0.001
Quantidade de exames histopatológicos	4678058.5	< 0.001
Antiviral	4195577.5	< 0.001
Cateter venoso central	4620861.5	< 0.001
Holter	4460778.5	< 0.001
Exames endoscópicos	4675980.0	< 0.001
Digoxina	4036938.0	< 0.001
Cintilografia	4596560.5	< 0.001
Betabloqueador	3963898.5	< 0.001
Tomografia	4481478.5	< 0.001
Diárias no serviço de Emergência na admissão T0	2480492.0	< 0.001
Quantidade de antimicrobianos	3754158.0	< 0.001
Antibióticos	3757923.0	< 0.001
Bloqueador do canal de calcio	4148967.5	< 0.001
Estatinas	3829690.0	< 0.001
Ano do procedimento 3	124289.5	< 0.001
Eletrofisiologia	4638403.0	< 0.001
IECA/BRA	3834341.0	< 0.001
Instalação de CEC	4713799.5	< 0.001
Bomba de infusão contínua	4082548.0	< 0.001
Outros procedimentos cirúrgicos	4614025.5	< 0.001
Tempo entre o P2 e P3	121424.0	< 0.001
Número de procedimentos em até 1 ano	6754413.5	< 0.001
Óbito hospitalar	6967140.0	< 0.001
Insulina	4107340.5	< 0.001
Anticonvulsivante	4148288.0	< 0.001
Transfusão de hemoderivados	4737640.5	< 0.001
Citologias	4755502.5	0.001
Diálise durante a admissão T0	6810550.0	0.001
Angio TC	4710936.5	0.002
Angio RM	4775332.5	0.003
Idade no Procedimento 2	899509.5	0.004

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

Variable	Statistic	p-value
Idade no momento do primeiro procedimento	7218506.5	0.006
Idade no Procedimento 1	7218506.5	0.006
Espirometria / Ergoespirometria	4770913.0	0.008
Intervenção coronária percutânea	4760306.0	0.012
Antiplaquetario EV	4234140.5	0.013
Teste de esforço	4758793.0	0.014
Tilt Test	4780797.5	0.014
Arteriografia	4791835.5	0.02
Cirurgia Toracica	4781540.0	0.021
Cardioversão/ Desfibrilação	4158206.0	0.033
PET-CT	4775117.0	0.035
Angioplastia	4789057.5	0.048
Dieta enteral	4164917.0	0.101
Interconsulta médica	4708945.0	0.107
Ventilação não invasiva	4817016.0	0.121
Ano da admissão T0	7019535.0	0.135
Ano do procedimento 1	7041442.0	0.153
Antihipertensivo	4219362.5	0.163
Marca-passo temporário	4157330.0	0.174
Flebografia	4770966.5	0.183
Intervenção cardiovascular em laboratório de hemodinâmica	4786111.5	0.191
Polissonografia	4795677.5	0.334
Trombolítico	4268704.0	0.411
Óbito intraoperatório	6853437.0	0.417
Drenagem de tórax e punção pericárdica ou pleural	4810277.5	0.476
Antiretroviral	4261728.5	0.494
Dieta parenteral	4180827.0	0.513
Número de Mudanças do tipo de DCEI	814086.0	0.545
Idade no Procedimento 3	102748.0	0.628
Cirurgia Cardiovascular	4817381.0	0.644
Angiografia	4798760.5	0.68
Aortografia	4798762.5	0.68
Número de procedimentos na admissão T0	6839453.5	0.746
Stent	4802405.0	0.804
Hipoglicemiante	4259804.5	0.883
Traqueostomia	4801161.0	0.886
Cavografia	4800696.5	0.932
Antiplaquetario VO	4264722.0	NaN
Hormonio tireoidiano	4264722.0	NaN
Broncodilator	4264722.0	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
Infarto do miocárdio prévio / Doença arterial coronariana	16.57	< 0.001
Insuficiência cardíaca	63.52	< 0.001
Número de procedimentos	93.41	< 0.001
Tipo de Procedimento 1	79.09	< 0.001
Tipo de Reoperação 1	91.86	< 0.001
Tipo de Dispositivo ao final do procedimento 1	88.64	< 0.001
Tipo de Reoperação 2	561.89	< 0.001
Tipo de Dispositivo ao final do procedimento 2	54.08	< 0.001
Óbito intraoperatório 2	43.77	< 0.001
Tipo de Reoperação 3	71.27	< 0.001
Tipo de Dispositivo ao final do procedimento 3	66.34	< 0.001
Óbito intraoperatório 3	64.03	< 0.001
Tipo de Reoperação 4	54.74	< 0.001
Tipo de Dispositivo ao final do procedimento 4	55.03	< 0.001
Óbito intraoperatório 4	47.66	< 0.001
Mudança do tipo de DCEI: entre o Procedimento 1 e Procedimento 2	43.30	< 0.001
Mudança do tipo de DCEI: entre o Procedimento 2 e Procedimento 3	64.08	< 0.001
Mudança do tipo de DCEI: entre o Procedimento 3 e Procedimento 4	49.09	< 0.001
UTI durante os episódios de hospitalização	136.15	< 0.001
Admissão em até 180 dias antes da T0	98.84	< 0.001
Readmissões pós-T0 com diárias de UTI	216.52	< 0.001
Desfecho principal da admissão T0	15.71	< 0.001
Readmissão cirúrgica em até 30 dias	2305.77	< 0.001
Readmissão cirúrgica entre 31 a 60 dias	1560.91	< 0.001
Readmissão cirúrgica em até 1 ano	13.02	< 0.001
Desfecho final do estudo	183.39	< 0.001
Causa do óbito	210.78	< 0.001
Ventilação mecânica / IOT	53.55	< 0.001
Escolaridade	27.50	0.001
Doença cardíaca	28.98	0.001
Classe funcional de IC	24.30	0.001
Doença cardíaca	15.37	0.003
Fibrilação / flutter atrial	9.59	0.004
Transplante cardíaco prévio	15.38	0.006
Díalise durante os episódios de hospitalização	9.51	0.007
Readmissão cirúrgica entre 61 a 180 dias	7.29	0.01
Parada cardíaca prévia/ Taquicardia ventricular instável	6.63	0.012
Valvopatias/ Prótese valvares	6.16	0.012

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

Variable	Statistic	p-value
Óbito intraoperatório 5	6.56	0.013
Diabetes mellitus	6.33	0.015
Tipo de Reoperação 7	16.36	0.015
Mudança do tipo de DCEI: entre o Procedimento 4 e Procedimento 5	8.43	0.022
Hemodiálise	6.48	0.025
Tipo de Reoperação 6	11.13	0.031
Tipo de Dispositivo ao final do procedimento 5	11.15	0.035
Óbito intraoperatório 6	4.54	0.038
Tipo de Reoperação 5	7.83	0.044
Tipo de Dispositivo ao final do procedimento 6	9.51	0.062
Readmissões pós-T0 com diálise	16.81	0.063
Mudança do tipo de DCEI: entre o Procedimento 5 e Procedimento 6	6.32	0.075
Mudança do tipo de DCEI: entre o Procedimento 6 e Procedimento 7	5.66	0.092
Óbito intraoperatório 7	2.82	0.107
Endocardite prévia	2.41	0.111
Estado de residência	36.85	0.14
Tipo de Dispositivo ao final do procedimento 7	5.28	0.172
Insuficiência renal crônica	1.58	0.229
Neoplasia em tratamento ou tratada recentemente	1.02	0.414
Doença pulmonar obstrutiva crônica	0.60	0.457
Acidente Vascular Cerebral/ Acidente isquêmico transitório prévios	0.44	0.555
Hipertensão arterial	0.35	0.585
Óbito intraoperatório 8	0.72	0.645
Mudança do tipo de DCEI: entre o Procedimento 7 e Procedimento 8	0.72	0.665
Tipo de Reoperação 8	0.72	0.734
Raça	2.06	0.88
Sexo	0.03	0.896
Óbito intraoperatório 1	0.42	> 0.999
Tipo de Dispositivo ao final do procedimento 8	0.72	> 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