

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
Óbito durante algum episódio de readmissão hospitalar	8380928	< 0.001
Óbito em até 180 dias após a alta T0	14603719	< 0.001
Óbito em até 2 anos após a alta T0	11798111	< 0.001
Óbito	0	< 0.001
Óbito em até 3 anos após a alta T0	10752625	< 0.001
Óbito em até 1 ano após a alta T0	13366340	< 0.001
Núm. de hospitalizações pós-procedimento	12044553	< 0.001
Idade no momento do primeiro procedimento	11499199	< 0.001
Idade no Procedimento 1	11499199	< 0.001
Número de atendimentos	12097466	< 0.001
Readmissão em até 1 ano	14160401	< 0.001
Ano da admissão T0	22299044	< 0.001
Ano do procedimento 1	22331294	< 0.001
Tempo de sobrevida	18038	< 0.001
Readmissão entre 61 a 180 dias	14945086	< 0.001
Óbito em até 30 dias após a alta T0	16317522	< 0.001
Tempo de seguimento total	21186196	< 0.001
Readmissão entre 31 a 60 dias	15864680	< 0.001
Readmissão em até 30 dias	16221644	< 0.001
Número da Admissão T0	15095022	< 0.001
Antibióticos	19878633	< 0.001
Ano do procedimento 2	1743602	< 0.001
Núm. de hospitalizações pré-procedimento	15377349	< 0.001
Tempo entre o P1 e P2	1646337	< 0.001
Interconsulta médica	18187512	< 0.001
Quantidade de classes medicamentosas utilizadas	18970707	< 0.001
Número de Mudanças do tipo de DCEI	1194294	< 0.001
ECG	18934074	< 0.001
Quantidade de exames por métodos gráficos	18932564	< 0.001
Quantidade de exames de análises clínicas	18720120	< 0.001
Exames laboratoriais	18717646	< 0.001
Idade no Procedimento 2	1086182	< 0.001
Radiografias	18656202	< 0.001
Tempo entre o P2 e P3	175274	< 0.001
Quantidade de exames diagnóstico por imagem	18633605	< 0.001
Bomba de infusão contínua	16899318	< 0.001

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

Variable	Statistic	p-value
IECA/BRA	18332036	< 0.001
Digoxina	16741096	< 0.001
Ano do procedimento 3	168569	< 0.001
Óbito hospitalar	17425496	< 0.001
Cirurgia Cardiovascular	17490476	< 0.001
Betabloqueador	17674238	< 0.001
Insuficiência cardíaca	16469466	< 0.001
Antagonista da Aldosterona	16517346	< 0.001
Anticoagulantes orais	16836384	< 0.001
Estatinas	17955966	< 0.001
Psicofármacos	17883743	< 0.001
Idade no Procedimento 3	113711	< 0.001
Antiarrítmicos	16718058	< 0.001
Antiplaquetário VO	17380672	< 0.001
Cintilografia	16961829	< 0.001
Dieta enteral	17124487	< 0.001
Quantidade de procedimentos invasivos	17603820	0.001
Ultrassom	16876117	0.002
Teste de esforço	17280128	0.003
Cavografia	17267208	0.003
Dieta parenteral	17156746	0.003
Antihipertensivo	17365966	0.005
Outros procedimentos cirúrgicos	17413529	0.005
Ecocardiograma	17651745	0.006
Instalação de CEC	17283889	0.007
Anticonvulsivante	17326790	0.019
Cateterismo	17404878	0.03
Insulina	17344028	0.049
Marca-passo temporário	17119086	0.053
Biopsias	17226078	0.066
Vasodilator	17468579	0.069
Hipoglicemiante	17302418	0.07
Antiplaquetário EV	17135332	0.073
Diárias no serviço de Emergência na admissão T0	17324270	0.113
Citologias	17138428	0.128
Trombolítico	17193904	0.149
Equipe Multiprofissional	17443944	0.149
Aortografia	17158170	0.153
Transplante cardíaco	17202077	0.167
Angioplastia	17162736	0.168
Angio TC	17252578	0.19
Antiretroviral	17191356	0.192
Cardioversão/ Desfibrilação	17155378	0.236
Intervenção coronária percutânea	17143223	0.244
Exames endoscópicos	17222014	0.279
Cirurgia Torácica	17197865	0.285
Tilt Test	17163272	0.308
Antifúngicos	17231480	0.308
Antiviral	17201940	0.31
Diurético	16996380	0.312
Broncodilatador	17231024	0.318
Ventilação não invasiva	17197528	0.335
Suporte cardiocirculatório	17197084	0.381

Table 2: Mann-Whitney Test (continued)

Variable	Statistic	p-value
Bloqueador do canal de calcio	17215937	0.414
Cateter venoso central	17225256	0.451
DVA	17072562	0.463
Ressonancia magnetica	17119928	0.465
Holter	17111845	0.495
Angio RM	17191122	0.511
Drenagem de tórax e punção pericárdica ou pleural	17165518	0.519
Quantidade de exames histopatológicos	17199912	0.58
Hormonio tireoidiano	17201492	0.605
Flebografia	17200446	0.641
Traqueostomia	17174942	0.663
Stent	17181164	0.664
PET-CT	17187617	0.75
Transfusão de hemoderivados	17189584	0.783
Arteriografia	17182066	0.793
Tomografia	17205920	0.8
Número de comorbidades	17136202	0.832
Culturas	17157048	0.848
Eletrofisiologia	17189274	0.885
Polissonografia	17178777	0.929
Angiografia	17180950	0.943
Intervenção cardiovascular em laboratório de hemodinâmica	17178766	0.961
Espirometria / Ergoespirometria	17180368	0.983

```
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	38.01	< 0.001
Escolaridade	427.01	< 0.001
Doença cardíaca	37.84	< 0.001
Doença cardíaca	18.02	< 0.001
Classe funcional de IC	173.94	< 0.001
Hipertensão arterial	20.14	< 0.001
Infarto do miocárdio prévio / Doença arterial coronariana	11.59	< 0.001
Fibrilação / flutter atrial	13.88	< 0.001
Valvopatias/ Prótese valvares	21.59	< 0.001
Insuficiência renal crônica	36.16	< 0.001
Número de procedimentos	64.91	< 0.001
Tipo de Reoperação 1	19.47	< 0.001
Tipo de Dispositivo ao final do procedimento 1	52.88	< 0.001
Óbito intraoperatório 1	37.06	< 0.001
Tipo de Reoperação 2	83.41	< 0.001
Tipo de Dispositivo ao final do procedimento 2	163.89	< 0.001
Óbito intraoperatório 2	51.28	< 0.001
Tipo de Dispositivo ao final do procedimento 3	64.71	< 0.001
Óbito intraoperatório 3	24.74	< 0.001
Mudança do tipo de DCEI: entre o Procedimento 1 e Procedimento 2	87.73	< 0.001
Diálise durante os episódios de hospitalização	34.41	< 0.001
UTI durante os episódios de hospitalização	121.39	< 0.001
Admissão em até 180 dias antes da T0	64.27	< 0.001
UTI durante a admissão T0	1758.63	< 0.001
Readmissões pós-T0 com diárias de UTI	158.05	< 0.001
Readmissões pós-T0 com diálise	54.58	< 0.001
Desfecho principal da admissão T0	34.07	< 0.001
Desfecho final do estudo	14323.58	< 0.001
Óbito intraoperatório	58.26	< 0.001
Causa do óbito	3392.08	< 0.001
Estado de residência	58.64	< 0.001
Hemodiálise	19.17	< 0.001
Tipo de Procedimento 1	10.36	< 0.001
Tipo de Reoperação 3	25.34	< 0.001
Tipo de Dispositivo ao final do procedimento 5	18.64	< 0.001
Tipo de Dispositivo ao final do procedimento 4	20.52	0.001
Tipo de Dispositivo ao final do procedimento 6	18.23	0.004
Sexo	8.26	0.007
Mudança do tipo de DCEI: entre o Procedimento 2 e Procedimento 3	10.32	0.007
Mudança do tipo de DCEI: entre o Procedimento 4 e Procedimento 5	11.30	0.007
Mudança do tipo de DCEI: entre o Procedimento 5 e Procedimento 6	9.88	0.011
Neoplasia em tratamento ou tratada recentemente	5.94	0.021
Insuficiência cardíaca	4.67	0.035
Tipo de Reoperação 4	7.53	0.056
Ventilação mecânica / IOT	3.45	0.065
Diabetes mellitus	3.05	0.087
Diálise durante a admissão T0	30.45	0.088
Óbito intraoperatório 6	3.27	0.098
Tipo de Dispositivo ao final do procedimento 7	6.85	0.15
Mudança do tipo de DCEI: entre o Procedimento 7 e Procedimento 8	5.34	0.163
Óbito intraoperatório 5	2.21	0.17
Tipo de Reoperação 6	4.37	0.235
Endocardite prévia	1.41	0.25

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

Variable	Statistic	p-value
Acidente Vascular Cerebral/ Acidente isquêmico transitório prévios	0.80	0.383
Tipo de Reoperação 5	3.04	0.398
Tipo de Dispositivo ao final do procedimento 8	0.74	0.42
Tipo de Reoperação 8	0.74	0.428
Óbito intraoperatório 8	0.74	0.428
Mudança do tipo de DCEI: entre o Procedimento 6 e Procedimento 7	1.25	0.454
Transplante cardíaco prévio	0.24	0.679
Mudança do tipo de DCEI: entre o Procedimento 3 e Procedimento 4	0.65	0.733
Parada cardíaca prévia/ Taquicardia ventricular instável	0.09	0.793
Óbito intraoperatório 4	0.06	0.842
Doença pulmonar obstrutiva crônica	0.02	0.923
Tipo de Reoperação 7	0.68	0.946
Óbito intraoperatório 7	0.00	> 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	0.19	> 0.999
Tipo de Dispositivo ao final do procedimento 10	0.19	> 0.999
Óbito intraoperatório 10	0.19	> 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	0.19	> 0.999