

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	161291.5	< 0.001
Óbito em até 3 anos após a alta T0	0.0	< 0.001
Óbito	128375.0	< 0.001
Tempo de sobrevida	130973.0	< 0.001
Núm. de hospitalizações pós-procedimento	438936.5	< 0.001
Número de atendimentos	442583.5	< 0.001
Número da Admissão T0	809770.5	< 0.001
Tempo de seguimento total	1626260.0	< 0.001
Digoxina	1039073.0	< 0.001
Núm. de hospitalizações pré-procedimento	865675.5	< 0.001
Antagonista da Aldosterona	941007.0	< 0.001
Readmissão em até 1 ano	975578.5	< 0.001
Insuficiência cardíaca	953229.5	< 0.001
Tempo entre o P1 e P2	91531.0	< 0.001
Diuretico	909077.5	< 0.001
Antiarrítmicos	992799.0	< 0.001
Quantidade de classes medicamentosas utilizadas	927589.5	< 0.001
Bomba de infusão contínua	1128588.5	< 0.001
Dieta parenteral	1166745.5	< 0.001
Readmissão entre 61 a 180 dias	1069964.0	< 0.001
IECA/BRA	983014.5	< 0.001
Anticoagulantes orais	1105852.5	< 0.001
Dieta enteral	1160492.5	< 0.001
Cardioversão/ Desfibrilação	1154917.5	< 0.001
DVA	1052030.5	< 0.001
Quantidade de exames por métodos gráficos	999468.0	0.001
Eletrofisiologia	1120957.5	0.001
ECG	1004466.5	0.001
Exames laboratoriais	1011207.0	0.002
Quantidade de exames de análises clínicas	1011696.5	0.002
Quantidade de procedimentos invasivos	1072150.0	0.002
Outros procedimentos cirúrgicos	1111692.0	0.002
Psicofármacos	1038464.5	0.003
Ultrassom	1109309.5	0.004
Equipe Multiprofissional	1044764.0	0.005
Quantidade de exames diagnóstico por imagem	1027375.5	0.005

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

Variable	Statistic	p-value
Ressonancia magnetica	1119188.5	0.007
Radiografias	1032979.5	0.007
Bloqueador do canal de calcio	1145584.0	0.011
Holter	1111497.0	0.011
Vasodilator	1073444.5	0.011
Tempo entre o P2 e P3	4981.0	0.014
Ecocardiograma	1070901.0	0.015
Traqueostomia	1167380.5	0.019
Ano da admissão T0	1298574.0	0.02
Ano do procedimento 1	1298666.0	0.021
Número de comorbidades	1055476.5	0.022
Angio TC	1146102.5	0.04
Readmissão em até 30 dias	1142308.5	0.05
Idade no Procedimento 2	65022.5	0.055
PET-CT	1163192.0	0.06
Transfusão de hemoderivados	1159508.5	0.065
Número de Mudanças do tipo de DCEI	44670.5	0.067
Marca-passo temporário	1159711.0	0.073
Diárias no serviço de Emergência na admissão T0	1135455.0	0.088
Insulina	1140302.0	0.092
Cateter venoso central	1149697.5	0.093
Cateterismo	1133779.0	0.117
Estatinas	1107313.5	0.119
Angiografia	1168170.0	0.121
Antibióticos	1093158.0	0.124
Hipoglicemiante	1147455.0	0.124
Anticonvulsivante	1151328.0	0.13
Hormonio tireoidiano	1158280.5	0.132
Readmissão entre 31 a 60 dias	1146330.5	0.163
Antihipertensivo	1196725.0	0.166
Tilt Test	1168486.0	0.177
Cirurgia Cardiovascular	1154575.5	0.217
Cintilografia	1154818.0	0.218
Angio RM	1168887.5	0.251
Interconsulta médica	1203589.5	0.282
Cavografia	1180813.0	0.332
Culturas	1147836.0	0.361
Quantidade de exames histopatológicos	1167197.0	0.394
Citologias	1169831.0	0.427
Antiplaquetario EV	1178206.0	0.444
Drenagem de tórax e punção pericárdica ou pleural	1177653.0	0.474
Antiviral	1177179.0	0.503
Flebografia	1167188.0	0.537
Exames endoscópicos	1168629.0	0.538
Espirometria / Ergoespirometria	1170540.5	0.551
Suporte cardiocirculatório	1176389.0	0.559
Ano do procedimento 2	56282.0	0.566
Teste de esforço	1168962.0	0.57
Transplante cardíaco	1176073.0	0.585
Intervenção cardiovascular em laboratório de hemodinâmica	1170789.5	0.592
Antiplaquetario VO	1181644.5	0.598
Ventilação não invasiva	1175915.0	0.599
Cirurgia Toracica	1175678.0	0.621
Aortografia	1175599.0	0.628

Table 2: Mann-Whitney Test (continued)

Variable	Statistic	p-value
Betabloqueador	1162563.5	0.66
Angioplastia	1174967.0	0.7
Polissonografia	1174967.0	0.7
Biopsias	1171593.0	0.715
Instalação de CEC	1170775.0	0.742
Idade no momento do primeiro procedimento	1156365.5	0.747
Idade no Procedimento 1	1156365.5	0.747
Trombolítico	1174493.0	0.771
Antiretroviral	1174493.0	0.771
Arteriografia	1174335.0	0.801
Idade no Procedimento 3	3115.0	0.804
Ano do procedimento 3	2699.0	0.805
Tomografia	1168407.0	0.823
Broncodilatador	1171789.0	0.866
Antifúngicos	1174433.0	0.958
Intervenção coronária percutânea	1173900.5	0.996
Óbito hospitalar	1173861.0	NaN
Óbito em até 30 dias após a alta T0	1173861.0	NaN
Óbito em até 180 dias após a alta T0	1173861.0	NaN
Óbito em até 1 ano após a alta T0	1173861.0	NaN
Óbito em até 2 anos após a alta T0	1173861.0	NaN
Stent	1173861.0	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') %>%
```

Table 3: Chi-squared test

Variable	Statistic	p-value
Insuficiência cardíaca	31.02	< 0.001
Tipo de Dispositivo ao final do procedimento 1	29.43	< 0.001
Tipo de Reoperação 2	42.79	< 0.001
Tipo de Dispositivo ao final do procedimento 2	32.47	< 0.001
Óbito intraoperatório 2	23.90	< 0.001
Mudança do tipo de DCEI: entre o Procedimento 1 e Procedimento 2	28.56	< 0.001
UTI durante os episódios de hospitalização	14.27	< 0.001
Admissão em até 180 dias antes da T0	38.58	< 0.001
Desfecho final do estudo	1185.20	< 0.001
Causa do óbito	176.54	< 0.001
Ventilação mecânica / IOT	17.89	< 0.001
Escolaridade	30.44	< 0.001
Classe funcional de IC	30.61	0.005
Óbito intraoperatório 3	9.37	0.005
UTI durante a admissão T0	2018.45	0.007
Doença cardíaca	24.54	0.009
Transplante cardíaco prévio	15.50	0.011
Sexo	5.20	0.022
Mudança do tipo de DCEI: entre o Procedimento 2 e Procedimento 3	9.43	0.022
Tipo de Reoperação 3	10.03	0.029
Tipo de Dispositivo ao final do procedimento 3	10.97	0.04
Hipertensão arterial	4.25	0.041
Número de procedimentos	24.75	0.051
Doença cardíaca	7.50	0.058
Raça	12.69	0.061
Readmissões pós-T0 com diárias de UTI	23.68	0.064
Óbito intraoperatório 4	3.43	0.068
Diálise durante a admissão T0	94.22	0.08
Valvopatias/ Prótese valvares	2.72	0.097
Infarto do miocárdio prévio / Doença arterial coronariana	2.82	0.114
Mudança do tipo de DCEI: entre o Procedimento 3 e Procedimento 4	3.45	0.139
Óbito intraoperatório 5	2.07	0.183
Tipo de Dispositivo ao final do procedimento 4	5.52	0.204
Tipo de Reoperação 4	3.52	0.227
Diabetes mellitus	1.42	0.26
Diálise durante os episódios de hospitalização	1.17	0.289
Mudança do tipo de DCEI: entre o Procedimento 4 e Procedimento 5	2.06	0.344
Insuficiência renal crônica	0.92	0.373
Tipo de Reoperação 5	2.06	0.472
Tipo de Dispositivo ao final do procedimento 5	2.06	0.556
Tipo de Procedimento 1	0.37	0.576
Neoplasia em tratamento ou tratada recentemente	0.74	0.641
Óbito intraoperatório 6	0.88	0.644
Mudança do tipo de DCEI: entre o Procedimento 5 e Procedimento 6	0.88	0.66
Fibrilação / flutter atrial	0.28	0.669
Tipo de Reoperação 1	0.67	0.884
Parada cardíaca prévia/ Taquicardia ventricular instável	0.04	0.896
Estado de residência	6.59	0.945
Endocardite prévia	0.09	> 0.999
Hemodiálise	0.14	> 0.999
Acidente Vascular Cerebral/ Acidente isquêmico transitório prévios	0.01	> 0.999

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

Variable	Statistic	p-value
Doença pulmonar obstrutiva crônica	0.00	> 0.999
Tipo de Reoperação 6	0.84	> 0.999
Tipo de Dispositivo ao final do procedimento 6	0.88	> 0.999
Tipo de Reoperação 7	0.34	> 0.999
Tipo de Dispositivo ao final do procedimento 7	0.33	> 0.999
Óbito intraoperatório 7	0.34	> 0.999
Tipo de Reoperação 8	0.13	> 0.999
Tipo de Dispositivo ao final do procedimento 8	0.13	> 0.999
Óbito intraoperatório 8	0.13	> 0.999
Tipo de Reoperação 9	0.05	> 0.999
Tipo de Dispositivo ao final do procedimento 9	0.05	> 0.999
Óbito intraoperatório 9	0.05	> 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	0.33	> 0.999
Mudança do tipo de DCEI: entre o Procedimento 7 e Procedimento 8	0.13	> 0.999
Mudança do tipo de DCEI: entre o Procedimento 8 e Procedimento 9	0.05	> 0.999
Mudança do tipo de DCEI: entre o Procedimento 9 e Procedimento 10	0.01	> 0.999
Readmissões pós-T0 com diálise	0.18	> 0.999
Óbito intraoperatório 1	NaN	NaN
Desfecho principal da admissão T0	NaN	NaN
Óbito intraoperatório	NaN	NaN