

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
Readmissão em até 30 dias	0.0	< 0.001
Readmissão entre 31 a 60 dias	92567.5	< 0.001
Readmissão entre 61 a 180 dias	270722.0	< 0.001
Readmissão em até 1 ano	429452.5	< 0.001
Núm. de hospitalizações pós-procedimento	1631771.5	< 0.001
Número de atendimentos	1896861.0	< 0.001
Óbito em até 30 dias após a alta T0	4292631.5	< 0.001
Óbito durante algum episódio de readmissão hospitalar	3586789.5	< 0.001
Óbito em até 180 dias após a alta T0	3799305.0	< 0.001
Tempo entre o P1 e P2	919146.0	< 0.001
Óbito	3723545.0	< 0.001
Quantidade de exames diagnóstico por imagem	3611246.0	< 0.001
Número da Admissão T0	3872501.0	< 0.001
Ecocardiograma	3869646.5	< 0.001
Radiografias	3703196.0	< 0.001
Quantidade de exames por métodos gráficos	3711517.5	< 0.001
ECG	3716746.5	< 0.001
Equipe Multiprofissional	3859669.5	< 0.001
Antiarrítmicos	4116497.5	< 0.001
Tempo de seguimento total	5616242.0	< 0.001
DVA	4063010.5	< 0.001
Tempo de sobrevida	245146.5	< 0.001
Exames laboratoriais	3824491.0	< 0.001
Quantidade de exames de análises clínicas	3824523.5	< 0.001
Diuretico	3941797.0	< 0.001
Vasodilator	4040784.0	< 0.001
Culturas	4206402.5	< 0.001
Quantidade de classes medicamentosas utilizadas	3862828.5	< 0.001
Antagonista da Aldosterona	4161725.0	< 0.001
Ultrassom	4319283.0	< 0.001
Quantidade de procedimentos invasivos	4196368.5	< 0.001
Núm. de hospitalizações pré-procedimento	4053751.5	< 0.001
Anticoagulantes orais	4416026.0	< 0.001
Psicofármacos	4037657.0	< 0.001
Suporte cardiocirculatório	4615242.0	< 0.001
Biopsias	4599825.0	< 0.001

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

Variable	Statistic	p-value
Ano do procedimento 2	674941.0	< 0.001
Cateterismo	4356943.0	< 0.001
Antifúngicos	4526116.5	< 0.001
Óbito em até 1 ano após a alta T0	3591007.5	< 0.001
Insuficiência cardíaca	4238250.5	< 0.001
Cintilografia	4489927.0	< 0.001
Ressonancia magnetica	4444415.0	< 0.001
Betabloqueador	4387508.5	< 0.001
Antibióticos	4086992.5	< 0.001
Digoxina	4471947.0	< 0.001
Tomografia	4403650.0	< 0.001
Quantidade de exames histopatológicos	4583623.0	< 0.001
Bloqueador do canal de calcio	4565977.0	< 0.001
Cateter venoso central	4524149.5	< 0.001
Número de comorbidades	4146637.5	< 0.001
Tempo entre o P2 e P3	91775.5	< 0.001
Diárias no serviço de Emergência na admissão T0	4457252.5	< 0.001
Holter	4436565.5	< 0.001
Transplante cardíaco	4642797.0	< 0.001
Exames endoscópicos	4591616.0	< 0.001
Bomba de infusão contínua	4578125.5	< 0.001
Óbito em até 2 anos após a alta T0	3421387.0	< 0.001
Broncodilator	4568338.5	< 0.001
Antiviral	4634352.5	< 0.001
Estatinas	4316718.0	< 0.001
Ano do procedimento 3	87991.5	< 0.001
Eletrofisiologia	4555974.5	< 0.001
IECA/BRA	4306877.0	< 0.001
Angio RM	4652049.0	< 0.001
Anticonvulsivante	4573139.0	< 0.001
Instalação de CEC	4614982.0	< 0.001
Óbito hospitalar	4761308.0	0.001
Outros procedimentos cirúrgicos	4546560.0	0.002
Insulina	4548089.5	0.002
Cardioversão/ Desfibrilação	4648863.0	0.002
Citologias	4643995.0	0.008
Angio TC	4608952.5	0.012
Intervenção coronária percutânea	4644718.5	0.024
Marca-passo temporário	4644933.0	0.025
Flebografia	4631090.5	0.028
PET-CT	4654611.5	0.032
Idade no Procedimento 2	592090.5	0.037
Antiplaquetario EV	4655530.0	0.043
Óbito em até 3 anos após a alta T0	3288355.5	0.05
Transfusão de hemoderivados	4645914.5	0.051
Intervenção cardiovascular em laboratório de hemodinâmica	4660565.5	0.078
Tilt Test	4667151.5	0.085
Dieta enteral	4667467.0	0.098
Angioplastia	4671223.0	0.103
Antihipertensivo	4631242.0	0.142
Teste de esforço	4656770.5	0.148
Arteriografia	4676199.0	0.196
Ventilação não invasiva	4693627.5	0.215

Table 2: Mann-Whitney Test (continued)

Variable	Statistic	p-value
Hormonio tireoidiano	4655986.0	0.237
Antiretroviral	4676507.0	0.25
Angiografia	4673650.0	0.296
Idade no momento do primeiro procedimento	4788502.5	0.34
Idade no Procedimento 1	4788502.5	0.34
Polissonografia	4687254.0	0.4
Drenagem de tórax e punção pericárdica ou pleural	4691385.5	0.409
Cirurgia Toracica	4675767.0	0.493
Trombolitico	4685129.5	0.511
Cavografia	4672058.5	0.523
Hipoglicemiante	4703181.0	0.544
Traqueostomia	4678329.0	0.559
Idade no Procedimento 3	73812.0	0.591
Dieta parenteral	4683915.5	0.6
Cirurgia Cardiovascular	4697189.0	0.64
Ano da admissão T0	4710592.5	0.656
Interconsulta médica	4656300.5	0.659
Ano do procedimento 1	4720909.0	0.726
Stent	4682094.5	0.843
Espirometria / Ergoespirometria	4683059.5	0.912
Número de Mudanças do tipo de DCEI	546417.0	0.922
Antiplaquetario VO	4679277.5	0.935
Aortografia	4681970.5	0.982

```

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
Insuficiência cardíaca	41.01	< 0.001
Número de procedimentos	62.85	< 0.001
Tipo de Procedimento 1	38.02	< 0.001
Tipo de Reoperação 1	50.60	< 0.001
Tipo de Dispositivo ao final do procedimento 1	51.76	< 0.001
Tipo de Reoperação 2	362.12	< 0.001
Tipo de Dispositivo ao final do procedimento 2	30.32	< 0.001
Óbito intraoperatório 2	21.24	< 0.001
Tipo de Reoperação 3	59.44	< 0.001
Tipo de Dispositivo ao final do procedimento 3	50.95	< 0.001
Óbito intraoperatório 3	47.56	< 0.001
Tipo de Reoperação 4	44.42	< 0.001
Tipo de Dispositivo ao final do procedimento 4	39.81	< 0.001
Óbito intraoperatório 4	35.43	< 0.001
Mudança do tipo de DCEI: entre o Procedimento 1 e Procedimento 2	20.89	< 0.001
Mudança do tipo de DCEI: entre o Procedimento 2 e Procedimento 3	47.24	< 0.001
Mudança do tipo de DCEI: entre o Procedimento 3 e Procedimento 4	35.67	< 0.001
UTI durante os episódios de hospitalização	105.97	< 0.001
Admissão em até 180 dias antes da T0	52.81	< 0.001
UTI durante a admissão T0	2759.45	< 0.001
Readmissões pós-T0 com diárias de UTI	149.96	< 0.001
Desfecho final do estudo	144.04	< 0.001
Causa do óbito	43.12	< 0.001
Ventilação mecânica / IOT	40.79	< 0.001
Desfecho principal da admissão T0	10.32	0.003
Transplante cardíaco prévio	18.47	0.004
Escolaridade	18.16	0.007
Infarto do miocárdio prévio / Doença arterial coronariana	7.32	0.009
Díálise durante os episódios de hospitalização	9.32	0.01
Doença cardíaca	11.60	0.011
Readmissões pós-T0 com diálise	26.24	0.026
Doença cardíaca	18.99	0.03
Endocardite prévia	4.58	0.037
Óbito intraoperatório 5	4.75	0.039
Hemodiálise	5.87	0.043
Classe funcional de IC	11.87	0.043
Mudança do tipo de DCEI: entre o Procedimento 4 e Procedimento 5	6.04	0.062
Tipo de Reoperação 5	7.38	0.065
Acidente Vascular Cerebral/ Acidente isquêmico transitório prévios	3.37	0.077
Tipo de Dispositivo ao final do procedimento 5	8.45	0.082
Valvopatias/ Prótese valvares	2.86	0.088
Insuficiência renal crônica	2.97	0.091
Díálise durante a admissão T0	45.10	0.095
Tipo de Reoperação 6	6.31	0.102
Mudança do tipo de DCEI: entre o Procedimento 5 e Procedimento 6	3.95	0.106
Óbito intraoperatório 6	2.82	0.128
Fibrilação / flutter atrial	2.42	0.131
Parada cardíaca prévia/ Taquicardia ventricular instável	2.03	0.162
Diabetes mellitus	1.70	0.207
Tipo de Dispositivo ao final do procedimento 6	4.77	0.211
Neoplasia em tratamento ou tratada recentemente	1.62	0.229
Estado de residência	33.65	0.241
Hipertensão arterial	0.72	0.418

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

Variable	Statistic	p-value
Sexo	0.39	0.526
Tipo de Dispositivo ao final do procedimento 7	1.25	0.689
Raça	1.54	0.923
Doença pulmonar obstrutiva crônica	0.01	> 0.999
Óbito intraoperatório 1	0.28	> 0.999
Tipo de Reoperação 7	0.71	> 0.999
Óbito intraoperatório 7	0.04	> 0.999
Tipo de Reoperação 8	0.47	> 0.999
Tipo de Dispositivo ao final do procedimento 8	0.47	> 0.999
Óbito intraoperatório 8	0.47	> 0.999
Tipo de Reoperação 9	0.20	> 0.999
Tipo de Dispositivo ao final do procedimento 9	0.20	> 0.999
Óbito intraoperatório 9	0.20	> 0.999
Tipo de Reoperação 10	0.04	> 0.999
Tipo de Dispositivo ao final do procedimento 10	0.04	> 0.999
Óbito intraoperatório 10	0.04	> 0.999
Mudança do tipo de DCEI: entre o Procedimento 6 e Procedimento 7	0.12	> 0.999
Mudança do tipo de DCEI: entre o Procedimento 7 e Procedimento 8	0.47	> 0.999
Mudança do tipo de DCEI: entre o Procedimento 8 e Procedimento 9	0.20	> 0.999
Mudança do tipo de DCEI: entre o Procedimento 9 e Procedimento 10	0.04	> 0.999
Óbito intraoperatório	0.43	> 0.999