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)
```

```r
# 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
```

```r
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
Número de atendimentos	4661386.0	< 0.001
Núm. de hospitalizações pós-procedimento	4006549.5	< 0.001
Readmissão em até 30 dias	9912782.5	< 0.001
Readmissão entre 31 a 60 dias	7776105.0	< 0.001
Readmissão entre 61 a 180 dias	3663876.5	< 0.001
Readmissão em até 1 ano	0.0	< 0.001
Óbito durante algum episódio de readmissão hospitalar	10818873.0	< 0.001
Tempo entre o P1 e P2	2740150.5	< 0.001
Óbito em até 180 dias após a alta T0	12496860.0	< 0.001
Óbito em até 1 ano após a alta T0	11472459.0	< 0.001
Óbito	11145632.0	< 0.001
Número da Admissão T0	11202873.5	< 0.001
Núm. de hospitalizações pré-procedimento	11717666.5	< 0.001
Antiarritmicos	12349895.5	< 0.001
Óbito em até 2 anos após a alta T0	10821897.0	< 0.001
DVA	12171862.0	< 0.001
Antagonista da Aldosterona	12318766.5	< 0.001
Insuficiência cardíaca	12218263.5	< 0.001
Quantidade de exames diagnóstico por imagem	11463725.5	< 0.001
Equipe Multiprofissional	11805316.0	< 0.001
Diuretico	11941290.5	< 0.001
Ultrassom	13009835.0	< 0.001
Quantidade de exames por métodos gráficos	11703278.5	< 0.001
Exames laboratoriais	11777192.0	< 0.001
Transplante cardíaco	13979005.5	< 0.001
Quantidade de exames de análises clínicas	11778710.0	< 0.001
Biopsias	13878840.5	< 0.001
Radiografias	11815745.5	< 0.001
ECG	11786388.0	< 0.001
Ecocardiograma	12283166.5	< 0.001
Quantidade de classes medicamentosas utilizadas	11872159.0	< 0.001
Holter	13085042.5	< 0.001
Óbito em até 30 dias após a alta T0	13640829.0	< 0.001
Quantidade de procedimentos invasivos	12779468.5	< 0.001
Número de comorbidades	12159184.5	< 0.001
Cateterismo	13158835.5	< 0.001

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

Variable	Statistic	p-value
Tempo de seguimento total	16200044.5	< 0.001
Quantidade de exames histopatológicos	13822470.0	< 0.001
Ressonancia magnetica	13387764.0	< 0.001
Ano do procedimento 2	2002392.0	< 0.001
Culturas	13079709.0	< 0.001
Anticoagulantes orais	13495401.0	< 0.001
Cateter venoso central	13638567.5	< 0.001
Tomografia	13280451.5	< 0.001
Psicofármacos	12653597.0	< 0.001
Cintilografia	13631812.5	< 0.001
Vasodilator	12863191.0	< 0.001
Bloqueador do canal de calcio	13809641.5	< 0.001
Antiviral	13996177.0	< 0.001
Digoxina	13585345.5	< 0.001
Ano do procedimento 3	245761.5	< 0.001
Estatinas	12864052.0	< 0.001
Exames endoscópicos	13887870.0	< 0.001
Eletrofisiologia	13701695.0	< 0.001
Tempo de sobrevida	537805.0	< 0.001
Bomba de infusão contínua	13878100.0	< 0.001
Suporte cardiocirculatório	14042213.5	< 0.001
Antifúngicos	13848330.5	< 0.001
Óbito em até 3 anos após a alta T0	10399157.5	< 0.001
Instalação de CEC	13926277.0	< 0.001
Tempo entre o P2 e P3	233941.5	< 0.001
Antibióticos	12900085.0	< 0.001
IECA/BRA	12977998.0	< 0.001
Idade no Procedimento 2	1864477.0	< 0.001
Diárias no serviço de Emergência na admissão T0	13633518.0	< 0.001
Betabloqueador	13584091.5	< 0.001
Óbito hospitalar	14430003.0	< 0.001
Outros procedimentos cirúrgicos	13728168.0	< 0.001
Intervenção coronária percutânea	14025735.0	< 0.001
Broncodiltador	13942464.5	< 0.001
Idade no momento do primeiro procedimento	15086741.0	< 0.001
Idade no Procedimento 1	15086741.0	< 0.001
Antiplaquetario EV	14061487.5	< 0.001
Cardioversão/ Desfibrilação	14079393.5	< 0.001
Transfusão de hemoderivados	14024134.5	< 0.001
Espirometria / Ergoespirometria	14078469.0	< 0.001
Angio RM	14103283.0	< 0.001
Citologias	14068657.5	< 0.001
Número de Mudanças do tipo de DCEI	1557238.5	< 0.001
Angio TC	13973357.0	< 0.001
Insulina	13879304.0	< 0.001
Angioplastia	14127210.5	< 0.001
Arteriografia	14141141.0	0.001
Anticonvulsivante	13984441.5	0.001
Intervenção cardiovascular em laboratório de hemodinâmica	14099645.0	0.002
Interconsulta médica	13901271.5	0.009
Dieta parenteral	14148149.5	0.016
Tilt Test	14132334.0	0.027
Flebografia	14077061.0	0.028

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

Variable	Statistic	p-value
Teste de esforço	14103088.0	0.039
Ano da admissão T0	14505441.0	0.044
Ano do procedimento 1	14544429.5	0.051
PET-CT	14122643.0	0.054
Ventilação não invasiva	14196539.5	0.058
Antiretroviral	14150171.5	0.061
Marca-passo temporário	14113927.0	0.073
Hormonio tireoidiano	14100194.5	0.087
Dieta enteral	14141357.0	0.114
Aortografia	14144314.5	0.132
Cirurgia Toracica	14142353.5	0.136
Polissonografia	14151250.0	0.219
Antihipertensivo	14099294.0	0.272
Idade no Procedimento 3	195054.5	0.279
Traqueostomia	14156234.5	0.388
Hipoglicemiante	14126339.0	0.527
Trombolitico	14160213.0	0.578
Cirurgia Cardiovascular	14195405.5	0.597
Stent	14166132.0	0.704
Drenagem de tórax e punção pericárdica ou pleural	14159576.0	0.784
Antiplaquetario VO	14152841.5	0.819
Angiografia	14167359.0	0.869
Cavografia	14162006.5	0.907

```
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
Sexo	23.71	< 0.001
Escolaridade	51.16	< 0.001
Doença cardíaca	103.19	< 0.001
Doença cardíaca	47.10	< 0.001
Classe funcional de IC	68.92	< 0.001
Infarto do miocárdio prévio / Doença arterial coronariana	33.29	< 0.001
Insuficiência cardíaca	196.28	< 0.001
Fibrilação / flutter atrial	14.89	< 0.001
Parada cardíaca prévia/ Taquicardia ventricular instável	20.27	< 0.001
Transplante cardíaco prévio	81.09	< 0.001
Valvopatias/ Prótese valvares	20.10	< 0.001
Insuficiência renal crônica	13.47	< 0.001
Número de procedimentos	199.74	< 0.001
Tipo de Procedimento 1	123.75	< 0.001
Tipo de Reoperação 1	153.24	< 0.001
Tipo de Dispositivo ao final do procedimento 1	294.05	< 0.001
Tipo de Reoperação 2	1067.46	< 0.001
Tipo de Dispositivo ao final do procedimento 2	203.24	< 0.001
Óbito intraoperatório 2	89.08	< 0.001
Tipo de Reoperação 3	111.24	< 0.001
Tipo de Dispositivo ao final do procedimento 3	142.94	< 0.001
Óbito intraoperatório 3	131.04	< 0.001
Tipo de Reoperação 4	147.55	< 0.001
Tipo de Dispositivo ao final do procedimento 4	141.78	< 0.001
Óbito intraoperatório 4	138.84	< 0.001
Tipo de Reoperação 5	40.35	< 0.001
Tipo de Dispositivo ao final do procedimento 5	37.72	< 0.001
Óbito intraoperatório 5	36.93	< 0.001
Óbito intraoperatório 6	17.82	< 0.001
Tipo de Dispositivo ao final do procedimento 7	33.07	< 0.001
Mudança do tipo de DCEI: entre o Procedimento 1 e Procedimento 2	93.20	< 0.001
Mudança do tipo de DCEI: entre o Procedimento 2 e Procedimento 3	131.96	< 0.001
Mudança do tipo de DCEI: entre o Procedimento 3 e Procedimento 4	138.95	< 0.001
Mudança do tipo de DCEI: entre o Procedimento 4 e Procedimento 5	37.78	< 0.001
Mudança do tipo de DCEI: entre o Procedimento 5 e Procedimento 6	23.29	< 0.001
Diálise durante os episódios de hospitalização	24.64	< 0.001
UTI durante os episódios de hospitalização	307.49	< 0.001
Admissão em até 180 dias antes da T0	249.76	< 0.001
UTI durante a admissão T0	2489.10	< 0.001
Readmissões pós-T0 com diárias de UTI	436.35	< 0.001
Desfecho principal da admissão T0	37.84	< 0.001
Desfecho final do estudo	476.98	< 0.001
Causa do óbito	100.77	< 0.001
Ventilação mecânica / IOT	91.62	< 0.001
Óbito intraoperatório 7	13.78	< 0.001
Readmissões pós-T0 com diálise	27.01	< 0.001
Estado de residência	61.67	0.002
Mudança do tipo de DCEI: entre o Procedimento 6 e Procedimento 7	17.67	0.003
Diabetes melittus	9.12	0.003
Tipo de Dispositivo ao final do procedimento 6	18.30	0.004
Tipo de Reoperação 6	16.91	0.006
Tipo de Reoperação 7	16.87	0.008
Hemodiálise	7.37	0.015

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

Variable	Statistic	p-value
Diálise durante a admissão T0	40.18	0.024
Raça	14.66	0.036
Endocardite prévia	3.83	0.053
Doença pulmonar obstrutiva crônica	2.95	0.104
Acidente Vascular Cerebral/ Acidente isquêmico transitório prévios	2.20	0.142
Mudança do tipo de DCEI: entre o Procedimento 7 e Procedimento 8	2.29	0.253
Neoplasia em tratamento ou tratada recentemente	1.50	0.261
Óbito intraoperatório	1.59	0.36
Tipo de Reoperação 8	1.67	0.376
Tipo de Dispositivo ao final do procedimento 8	1.67	0.38
Óbito intraoperatório 8	1.67	0.39
Óbito intraoperatório 1	1.01	0.594
Hipertensão arterial	0.13	0.746
Tipo de Reoperação 9	0.25	> 0.999
Tipo de Dispositivo ao final do procedimento 9	0.25	> 0.999
Óbito intraoperatório 9	0.25	> 0.999
Tipo de Reoperação 10	0.14	> 0.999
Tipo de Dispositivo ao final do procedimento 10	0.14	> 0.999
Óbito intraoperatório 10	0.14	> 0.999
Mudança do tipo de DCEI: entre o Procedimento 8 e Procedimento 9	0.25	> 0.999
Mudança do tipo de DCEI: entre o Procedimento 9 e Procedimento 10	0.14	> 0.999