

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){
  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(.): 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 > 0.8) %>%
  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	0.85
Idade no momento do primeiro procedimento	Idade no Procedimento 2	0.85
Ano do procedimento 1	Ano da admissão T0	0.85
Idade no Procedimento 1	Idade no momento do primeiro procedimento	0.85
Idade no Procedimento 1	Idade no Procedimento 2	0.85
Idade no Procedimento 2	Idade no momento do primeiro procedimento	0.85
Idade no Procedimento 2	Idade no Procedimento 1	0.85
Número de atendimentos	Núm. de hospitalizações pós-procedimento	0.85
Número da Admissão T0	Núm. de hospitalizações pré-procedimento	0.85
Núm. de hospitalizações pós-procedimento	Número de atendimentos	0.85
Núm. de hospitalizações pré-procedimento	Número da Admissão T0	0.85
Ano da admissão T0	Ano do procedimento 1	0.85
Readmissão em até 30 dias	Readmissão entre 31 a 60 dias	0.85
Readmissão entre 31 a 60 dias	Readmissão em até 30 dias	0.85
Readmissão entre 61 a 180 dias	Readmissão em até 1 ano	0.85
Readmissão em até 1 ano	Readmissão entre 61 a 180 dias	0.85
Óbito durante algum episódio de readmissão hospitalar	Óbito	0.85
Óbito	Óbito durante algum episódio de readmissão hospitalar	0.85
DVA	Quantidade de medicamentos de ação cardiovascular	0.85
Diuretico	Quantidade de medicamentos de ação cardiovascular	0.85
Vasodilator	Quantidade de medicamentos de ação cardiovascular	0.85
Antibióticos	Quantidade de antimicrobianos	0.85
Quantidade de classes medicamentosas utilizadas	Quantidade de classes medicamentosas de ação cardiovascular	0.85
Quantidade de classes medicamentosas de ação cardiovascular	Quantidade de classes medicamentosas utilizadas	0.85

Table 1: Pearson Correlation (*continued*)

row	column	cor
Quantidade de medicamentos de ação cardiovascular	DVA	
Quantidade de medicamentos de ação cardiovascular	Diuretico	
Quantidade de medicamentos de ação cardiovascular	Vasodilator	
Quantidade de antimicrobianos	Antibióticos	
Stent	Angioplastia	
Stent	Interconsulta médica	
Angioplastia	Stent	
Interconsulta médica	Stent	
Equipe Multiprofissional	Radiografias	
Equipe Multiprofissional	Quantidade de exames diagnóstico por imagem	
ECG	Quantidade de exames por métodos gráficos	
ECG	Exames laboratoriais	
ECG	Quantidade de exames de análises clínicas	
ECG	Radiografias	
ECG	Quantidade de exames diagnóstico por imagem	
Quantidade de exames por métodos gráficos	ECG	
Quantidade de exames por métodos gráficos	Exames laboratoriais	
Quantidade de exames por métodos gráficos	Quantidade de exames de análises clínicas	
Quantidade de exames por métodos gráficos	Radiografias	
Quantidade de exames por métodos gráficos	Quantidade de exames diagnóstico por imagem	
Exames laboratoriais	ECG	
Exames laboratoriais	Quantidade de exames por métodos gráficos	
Exames laboratoriais	Quantidade de exames de análises clínicas	
Exames laboratoriais	Radiografias	
Exames laboratoriais	Quantidade de exames diagnóstico por imagem	
Quantidade de exames de análises clínicas	ECG	
Quantidade de exames de análises clínicas	Quantidade de exames por métodos gráficos	
Quantidade de exames de análises clínicas	Exames laboratoriais	
Quantidade de exames de análises clínicas	Radiografias	
Quantidade de exames de análises clínicas	Quantidade de exames diagnóstico por imagem	
Biopsias	Quantidade de exames histopatológicos	
Quantidade de exames histopatológicos	Biopsias	
Radiografias	Equipe Multiprofissional	
Radiografias	ECG	
Radiografias	Quantidade de exames por métodos gráficos	
Radiografias	Exames laboratoriais	
Radiografias	Quantidade de exames de análises clínicas	
Radiografias	Quantidade de exames diagnóstico por imagem	
Quantidade de exames diagnóstico por imagem	Equipe Multiprofissional	
Quantidade de exames diagnóstico por imagem	ECG	
Quantidade de exames diagnóstico por imagem	Quantidade de exames por métodos gráficos	
Quantidade de exames diagnóstico por imagem	Exames laboratoriais	
Quantidade de exames diagnóstico por imagem	Quantidade de exames de análises clínicas	
Quantidade de exames diagnóstico por imagem	Radiografias	

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	3083964.5	< 0.001
Núm. de hospitalizações pós-procedimento	3006306.5	< 0.001
Óbito durante algum episódio de readmissão hospitalar	0.0	< 0.001
Óbito	974088.0	< 0.001
Readmissão em até 1 ano	6262304.0	< 0.001
Readmissão entre 61 a 180 dias	7115096.0	< 0.001
Readmissão entre 31 a 60 dias	8105032.0	< 0.001
Readmissão em até 30 dias	8508376.0	< 0.001
Número da Admissão T0	7215350.5	< 0.001
Quantidade de classes medicamentosas utilizadas	2257196.0	< 0.001
Núm. de hospitalizações pré-procedimento	7410466.5	< 0.001
Quantidade de classes medicamentosas de ação cardiovascular	1945578.5	< 0.001
Tempo de sobrevida	270577.0	< 0.001
Ano da admissão T0	12022582.0	< 0.001
Ano do procedimento 1	12037696.0	< 0.001
Insuficiência cardíaca	4070322.0	< 0.001
Antagonista da Aldosterona	4184940.5	< 0.001
Antiarrítmicos	4393536.0	< 0.001
Diurético	4083932.0	< 0.001
Quantidade de medicamentos de ação cardiovascular	4022339.0	< 0.001
Digoxina	4783904.0	< 0.001
Tempo entre o P1 e P2	1121732.5	< 0.001
Tempo de seguimento total	11244076.0	< 0.001

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

Variable	Statistic	p-value
Ano do procedimento 2	1105696.0	< 0.001
Número de Mudanças do tipo de DCEI	714992.0	< 0.001
Anticoagulantes orais	4870796.0	< 0.001
DVA	4523258.0	< 0.001
Ultrassom	5670569.5	< 0.001
Bomba de infusão contínua	4963057.0	< 0.001
Equipe Multiprofissional	5407599.5	< 0.001
Tempo entre o P2 e P3	128921.5	< 0.001
Quantidade de exames diagnóstico por imagem	5387152.5	< 0.001
Quantidade de exames por métodos gráficos	5404498.5	< 0.001
IECA/BRA	4590970.0	< 0.001
Holter	5777937.5	< 0.001
Ecocardiograma	5523814.5	< 0.001
Cintilografia	5921737.0	< 0.001
ECG	5438056.5	< 0.001
Ressonancia magnetica	5880330.0	< 0.001
Ano do procedimento 3	122405.5	< 0.001
Radiografias	5553748.5	< 0.001
Vasodilator	4820794.0	< 0.001
Psicofármacos	4781603.0	< 0.001
Exames laboratoriais	5594256.5	< 0.001
Quantidade de exames de análises clínicas	5595883.5	< 0.001
Interconsulta médica	6461332.5	< 0.001
Número de procedimentos em até 1 ano	9484317.0	< 0.001
Estatinas	4871403.0	< 0.001
Eletrofisiologia	6006000.0	< 0.001
Diárias no serviço de Emergência na admissão T0	2754501.0	< 0.001
Quantidade de procedimentos invasivos	5861488.0	< 0.001
Óbito hospitalar	9736216.0	< 0.001
Culturas	5891335.5	< 0.001
Dieta enteral	5126507.0	< 0.001
Idade no momento do primeiro procedimento	10190528.5	< 0.001
Idade no Procedimento 1	10190528.5	< 0.001
Dieta parenteral	5146368.0	< 0.001
Citologias	6112230.0	< 0.001
Óbito intraoperatório	9572564.0	< 0.001
Cardioversão/ Desfibrilação	5127454.0	0.001
Angioplastia	6144292.5	0.001
Número de comorbidades	9108062.0	0.001
Tilt Test	6137587.5	0.002
Tomografia	5977787.5	0.002
Antiplaquetario EV	5211523.0	0.002
Cateterismo	5977309.0	0.002
Número de procedimentos em até 180 dias	9514640.0	0.003
Antihipertensivo	5356533.5	0.005
Idade no Procedimento 2	914505.5	0.005
Cirurgia Cardiovascular	6273864.0	0.005
Marca-passo temporário	5113841.5	0.005
Antifúngicos	5183285.0	0.01
Número de procedimentos em até 30 dias	9537415.5	0.011
Drenagem de tórax e punção pericárdica ou pleural	6136845.0	0.016
Bloqueador do canal de calcio	5197141.5	0.016
Cavografia	6205470.5	0.034

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

Variable	Statistic	p-value
Quantidade de antimicrobianos	5042530.5	0.038
Número de procedimentos em até 60 dias	9546655.5	0.042
Cateter venoso central	6105257.5	0.043
Antibióticos	5049643.0	0.044
Angio TC	6105699.0	0.058
Quantidade de exames histopatológicos	6128770.5	0.06
Espirometria / Ergoespirometria	6144054.0	0.066
Intervenção coronária percutânea	6133830.5	0.067
Polissonografia	6157559.0	0.146
Outros procedimentos cirúrgicos	6098098.5	0.156
Insulina	5190812.5	0.164
Flebografia	6131848.5	0.167
PET-CT	6149567.0	0.193
Suporte cardiocirculatório	6154823.5	0.216
Ventilação não invasiva	6181392.0	0.235
Exames endoscópicos	6142606.0	0.261
Intervenção cardiovascular em laboratório de hemodinâmica	6153259.5	0.273
Teste de esforço	6186948.0	0.35
Trombolítico	5259330.0	0.354
Antiretroviral	5258424.0	0.402
Angiografia	6161584.0	0.448
Antiviral	5246236.5	0.483
Transfusão de hemoderivados	6154194.5	0.497
Betabloqueador	5289453.0	0.528
Transplante cardíaco	6172826.5	0.589
Arteriografia	6165778.0	0.601
Instalação de CEC	6179560.5	0.62
Anticonvulsivante	5269168.0	0.666
Angio RM	6172238.5	0.706
Cirurgia Toracica	6171729.5	0.74
Stent	6168882.0	0.775
Idade no Procedimento 3	93415.0	0.838
Traqueostomia	6169296.0	0.893
Número de procedimentos na admissão T0	9601827.5	0.894
Hipoglicemiante	5249887.0	0.904
Aortografia	6168214.5	0.985
Biopsias	6168497.0	0.993
Antiplaquetario VO	5254347.0	NaN
Hormonio tireoidiano	5254347.0	NaN
Broncodilator	5254347.0	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
Sexo	23.94	< 0.001
Raça	44.24	< 0.001
Escolaridade	181.61	< 0.001
Doença cardíaca	63.16	< 0.001
Doença cardíaca	21.37	< 0.001
Classe funcional de IC	139.41	< 0.001
Hipertensão arterial	54.08	< 0.001
Insuficiência cardíaca	88.46	< 0.001
Tipo de Procedimento 1	17.85	< 0.001
Tipo de Reoperação 1	20.81	< 0.001
Tipo de Dispositivo ao final do procedimento 1	260.34	< 0.001
Tipo de Reoperação 2	65.54	< 0.001
Tipo de Dispositivo ao final do procedimento 2	165.91	< 0.001
Tipo de Reoperação 3	45.53	< 0.001
Tipo de Dispositivo ao final do procedimento 3	63.94	< 0.001
Óbito intraoperatório 3	48.13	< 0.001
Mudança do tipo de DCEI: entre o Procedimento 1 e Procedimento 2	66.74	< 0.001
Diálise durante os episódios de hospitalização	71.61	< 0.001
UTI durante os episódios de hospitalização	213.78	< 0.001
Admissão em até 180 dias antes da T0	168.35	< 0.001
UTI durante a admissão T0	2318.58	< 0.001
Readmissões pós-T0 com diárias de UTI	318.93	< 0.001
Readmissões pós-T0 com diálise	125.34	< 0.001
Desfecho principal da admissão T0	23.23	< 0.001
Desfecho final do estudo	6709.98	< 0.001
Causa do óbito	7202.55	< 0.001
Ventilação mecânica / IOT	18.00	< 0.001
Mudança do tipo de DCEI: entre o Procedimento 2 e Procedimento 3	16.66	< 0.001
Insuficiência renal crônica	11.45	0.001
Tipo de Dispositivo ao final do procedimento 4	18.19	0.002
Tipo de Reoperação 4	15.99	0.004
Parada cardíaca prévia/ Taquicardia ventricular instável	9.02	0.005
Tipo de Dispositivo ao final do procedimento 6	19.48	0.005
Readmissão cirúrgica em até 1 ano	8.76	0.005
Mudança do tipo de DCEI: entre o Procedimento 5 e Procedimento 6	12.91	0.006
Valvopatias/ Prótese valvares	7.32	0.008
Fibrilação / flutter atrial	7.27	0.009
Readmissão cirúrgica em até 30 dias	7.72	0.009

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

Variable	Statistic	p-value
Endocardite prévia	7.52	0.01
Estado de residência	54.33	0.012
Diálise durante a admissão T0	49.02	0.017
Readmissão cirúrgica entre 61 a 180 dias	5.62	0.019
Óbito intraoperatório 4	5.21	0.022
Readmissão cirúrgica entre 31 a 60 dias	5.98	0.023
Mudança do tipo de DCEI: entre o Procedimento 4 e Procedimento 5	8.01	0.025
Mudança do tipo de DCEI: entre o Procedimento 3 e Procedimento 4	7.12	0.035
Óbito intraoperatório 6	4.86	0.041
Número de procedimentos	19.65	0.043
Tipo de Reoperação 6	8.65	0.044
Infarto do miocárdio prévio / Doença arterial coronariana	3.88	0.065
Mudança do tipo de DCEI: entre o Procedimento 7 e Procedimento 8	11.30	0.091
Tipo de Dispositivo ao final do procedimento 7	11.65	0.105
Óbito intraoperatório 2	2.31	0.119
Doença pulmonar obstrutiva crônica	2.10	0.174
Acidente Vascular Cerebral/ Acidente isquêmico transitório prévios	1.82	0.188
Diabetes melittus	1.60	0.217
Tipo de Dispositivo ao final do procedimento 5	5.50	0.225
Tipo de Dispositivo ao final do procedimento 8	1.17	0.259
Tipo de Reoperação 8	1.17	0.261
Óbito intraoperatório 8	1.17	0.278
Óbito intraoperatório 5	1.35	0.289
Tipo de Reoperação 9	0.94	0.34
Tipo de Dispositivo ao final do procedimento 9	0.94	0.34
Mudança do tipo de DCEI: entre o Procedimento 8 e Procedimento 9	0.94	0.353
Mudança do tipo de DCEI: entre o Procedimento 6 e Procedimento 7	2.58	0.363
Óbito intraoperatório 9	0.94	0.383
Hemodiálise	0.89	0.427
Óbito intraoperatório 7	0.82	0.506
Tipo de Reoperação 5	1.75	0.621
Tipo de Reoperação 7	1.78	0.622
Transplante cardíaco prévio	0.91	0.642
Óbito intraoperatório 1	0.62	0.667
Neoplasia em tratamento ou tratada recentemente	0.05	0.87
Tipo de Reoperação 10	0.09	> 0.999
Tipo de Dispositivo ao final do procedimento 10	0.09	> 0.999
Óbito intraoperatório 10	0.09	> 0.999
Mudança do tipo de DCEI: entre o Procedimento 9 e Procedimento 10	0.09	> 0.999