

# 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
Diálise durante a admissão T0	Stent	0.85
Diálise durante a admissão T0	Angioplastia	0.85
Diálise durante a admissão T0	Interconsulta médica	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

Table 1: Pearson Correlation (*continued*)

row	column	cor
Antibióticos	Quantidade de antimicrobianos	
Quantidade de classes medicamentosas utilizadas	Quantidade de classes medicamentosas de ação cardiovascular	
Quantidade de classes medicamentosas de ação cardiovascular	Quantidade de classes medicamentosas utilizadas	
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	Diálise durante a admissão T0	
Stent	Angioplastia	
Stent	Interconsulta médica	
Angioplastia	Diálise durante a admissão T0	
Angioplastia	Stent	
Interconsulta médica	Diálise durante a admissão T0	
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))
}

## Can't calculate Wilcox test for variable survival_time_death

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	9879822.0	< 0.001
Óbito	0.0	< 0.001
Óbito hospitalar	16770568.0	< 0.001
Tempo de seguimento total	24282062.0	< 0.001
Idade no momento do primeiro procedimento	12855338.0	< 0.001
Idade no Procedimento 1	12855338.0	< 0.001
Ano da admissão T0	23562214.5	< 0.001
Ano do procedimento 1	23591798.5	< 0.001
Número de atendimentos	13753834.5	< 0.001
Núm. de hospitalizações pós-procedimento	14194259.0	< 0.001
Readmissão em até 1 ano	15741610.0	< 0.001
Quantidade de classes medicamentosas utilizadas	5099865.0	< 0.001
Readmissão entre 61 a 180 dias	16460105.0	< 0.001
UTI durante a admissão T0	15914550.0	< 0.001
Quantidade de classes medicamentosas de ação cardiovascular	4287569.5	< 0.001
Diuretico	8146445.0	< 0.001

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

Variable	Statistic	p-value
Número da Admissão T0	16094311.0	< 0.001
DVA	8497869.0	< 0.001
Readmissão entre 31 a 60 dias	17310984.0	< 0.001
Ultrassom	10748193.5	< 0.001
Ano do procedimento 2	1787659.5	< 0.001
Antifúngicos	9564242.5	< 0.001
Antiarrítmicos	8898540.0	< 0.001
Readmissão em até 30 dias	17624191.0	< 0.001
Núm. de hospitalizações pré-procedimento	16397166.0	< 0.001
Culturas	10625433.0	< 0.001
Antagonista da Aldosterona	8850457.0	< 0.001
Quantidade de medicamentos de ação cardiovascular	8401482.5	< 0.001
Insuficiência cardíaca	8803530.0	< 0.001
Tempo entre o P1 e P2	1726581.5	< 0.001
Equipe Multiprofissional	10276737.5	< 0.001
Bomba de infusão contínua	9534828.5	< 0.001
Citologias	11551407.0	< 0.001
Diálise durante a admissão T0	18326742.5	< 0.001
Tomografia	11049785.5	< 0.001
Digoxina	9577342.5	< 0.001
Ecocardiograma	10625542.5	< 0.001
Número de Mudanças do tipo de DCEI	1230274.5	< 0.001
Quantidade de exames diagnóstico por imagem	10602089.5	< 0.001
Anticoagulantes orais	9643846.5	< 0.001
Tempo entre o P2 e P3	178782.5	< 0.001
Óbito intraoperatório	18430431.0	< 0.001
Vasodilator	9212176.5	< 0.001
Psicofármacos	9119547.5	< 0.001
Idade no Procedimento 2	1127693.0	< 0.001
Quantidade de exames histopatológicos	11567985.5	< 0.001
Exames endoscópicos	11550384.0	< 0.001
Radiografias	10738697.0	< 0.001
Dieta enteral	9797926.0	< 0.001
Ano do procedimento 3	170686.0	< 0.001
Transfusão de hemoderivados	11597730.5	< 0.001
Cintilografia	11464017.0	< 0.001
Quantidade de exames por métodos gráficos	10861742.0	< 0.001
ECG	10863787.5	< 0.001
Interconsulta médica	12258650.5	< 0.001
Anticonvulsivante	9802943.5	< 0.001
Dieta parenteral	9849032.5	< 0.001
Exames laboratoriais	10958172.0	< 0.001
Quantidade de exames de análises clínicas	10959417.5	< 0.001
Insulina	9737943.5	< 0.001
Quantidade de procedimentos invasivos	11286772.5	< 0.001
Outros procedimentos cirúrgicos	11468587.0	< 0.001
Diárias no serviço de Emergência na admissão T0	5448459.0	< 0.001
Antiplaquetário EV	9976446.0	< 0.001
Holter	11445508.0	< 0.001
Idade no Procedimento 3	114288.5	< 0.001
Cateter venoso central	11615831.5	< 0.001
Suporte cardiocirculatório	11713933.5	< 0.001
Quantidade de antimicrobianos	9580480.5	< 0.001

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

Variable	Statistic	p-value
Drenagem de tórax e punção pericárdica ou pleural	11705326.5	< 0.001
Antibióticos	9591131.0	0.001
Marca-passo temporário	9807109.5	0.002
Teste de esforço	11848810.0	0.002
Ventilação não invasiva	11720580.0	0.003
Intervenção coronária percutânea	11689009.5	0.003
Cardioversão/ Desfibrilação	9848289.0	0.012
Número de comorbidades	17976091.5	0.013
Antiviral	10013832.5	0.014
PET-CT	11718079.5	0.017
Stent	11760085.0	0.022
Cirurgia Toracica	11734178.0	0.024
Cateterismo	11589096.0	0.037
Aortografia	11740068.5	0.038
Angioplastia	11744871.5	0.043
Traqueostomia	11748326.0	0.065
Betabloqueador	10194551.0	0.066
Cirurgia Cardiovascular	11854139.0	0.089
Ressonancia magnetica	11654969.0	0.101
Angio TC	11690815.5	0.101
Número de procedimentos em até 1 ano	18451024.5	0.154
Polissonografia	11751520.0	0.169
Hipoglicemiante	10118417.5	0.201
Número de procedimentos na admissão T0	18445426.0	0.205
IECA/BRA	9882219.0	0.218
Eletrofisiologia	11702459.0	0.237
Trombolitico	10044636.0	0.262
Estatinas	9915288.5	0.267
Bloqueador do canal de calcio	10017052.5	0.276
Cavografia	11791660.5	0.282
Tilt Test	11751979.5	0.311
Flebografia	11729975.5	0.328
Transplante cardíaco	11775621.5	0.381
Arteriografia	11760845.5	0.483
Número de procedimentos em até 30 dias	18482500.5	0.528
Instalação de CEC	11749646.0	0.607
Intervenção cardiovascular em laboratório de hemodinâmica	11758246.5	0.697
Antiretroviral	10055253.5	0.735
Antihipertensivo	10066120.0	0.793
Angio RM	11768390.0	0.847
Biopsias	11762556.0	0.875
Angiografia	11766587.5	0.94
Número de procedimentos em até 180 dias	18505867.5	0.945
Número de procedimentos em até 60 dias	18502137.0	0.976
Espirometria / Ergoespirometria	11766170.0	0.978
Tempo de sobrevida	NaN	NaN
Antiplaquetario VO	10052975.0	NaN
Hormonio tireoidiano	10052975.0	NaN
Broncodilator	10052975.0	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
Raça	76.28	< 0.001
Escolaridade	475.32	< 0.001
Classe funcional de IC	212.53	< 0.001
Hipertensão arterial	22.74	< 0.001
Infarto do miocárdio prévio / Doença arterial coronariana	15.60	< 0.001
Insuficiência cardíaca	31.69	< 0.001
Valvopatias/ Prótese valvares	24.16	< 0.001
Insuficiência renal crônica	58.48	< 0.001
Hemodiálise	26.49	< 0.001
Número de procedimentos	101.07	< 0.001
Tipo de Dispositivo ao final do procedimento 1	77.18	< 0.001
Óbito intraoperatório 1	33.10	< 0.001
Tipo de Reoperação 2	130.88	< 0.001
Tipo de Dispositivo ao final do procedimento 2	191.32	< 0.001
Óbito intraoperatório 2	89.54	< 0.001
Tipo de Reoperação 3	30.59	< 0.001
Tipo de Dispositivo ao final do procedimento 3	65.63	< 0.001
Óbito intraoperatório 3	29.60	< 0.001
Mudança do tipo de DCEI: entre o Procedimento 1 e Procedimento 2	123.72	< 0.001
Díálise durante os episódios de hospitalização	126.98	< 0.001
UTI durante os episódios de hospitalização	309.16	< 0.001
Admissão em até 180 dias antes da T0	98.42	< 0.001
Readmissões pós-T0 com diárias de UTI	126.31	< 0.001
Readmissões pós-T0 com diálise	47.93	< 0.001
Desfecho principal da admissão T0	1239.51	< 0.001
Desfecho final do estudo	16024.00	< 0.001
Causa do óbito	16024.00	< 0.001
Doença cardíaca	35.50	< 0.001
Doença cardíaca	20.50	< 0.001

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

Variable	Statistic	p-value
Tipo de Dispositivo ao final do procedimento 4	19.45	< 0.001
Mudança do tipo de DCEI: entre o Procedimento 2 e Procedimento 3	16.00	0.001
Fibrilação / flutter atrial	10.28	0.002
Tipo de Dispositivo ao final do procedimento 6	17.52	0.002
Sexo	9.46	0.002
Estado de residência	59.70	0.002
Tipo de Dispositivo ao final do procedimento 5	15.93	0.005
Ventilação mecânica / IOT	7.71	0.006
Neoplasia em tratamento ou tratada recentemente	5.98	0.017
Mudança do tipo de DCEI: entre o Procedimento 4 e Procedimento 5	9.13	0.019
Mudança do tipo de DCEI: entre o Procedimento 5 e Procedimento 6	8.06	0.021
Tipo de Reoperação 4	7.69	0.045
Endocardite prévia	3.17	0.087
Tipo de Procedimento 1	2.26	0.136
Tipo de Dispositivo ao final do procedimento 8	6.17	0.14
Tipo de Dispositivo ao final do procedimento 7	6.61	0.178
Mudança do tipo de DCEI: entre o Procedimento 7 e Procedimento 8	4.73	0.206
Óbito intraoperatório 6	2.03	0.207
Tipo de Reoperação 1	3.97	0.261
Tipo de Reoperação 6	2.96	0.399
Óbito intraoperatório 5	0.73	0.441
Readmissão cirúrgica em até 30 dias	0.48	0.473
Diabetes mellitus	0.53	0.495
Acidente Vascular Cerebral/ Acidente isquêmico transitório prévios	0.45	0.503
Readmissão cirúrgica entre 31 a 60 dias	0.57	0.506
Mudança do tipo de DCEI: entre o Procedimento 6 e Procedimento 7	1.41	0.507
Óbito intraoperatório 4	0.48	0.513
Mudança do tipo de DCEI: entre o Procedimento 3 e Procedimento 4	1.02	0.591
Transplante cardíaco prévio	0.28	0.69
Óbito intraoperatório 8	0.47	0.692
Tipo de Reoperação 5	1.46	0.701
Readmissão cirúrgica em até 1 ano	0.20	0.715
Doença pulmonar obstrutiva crônica	0.10	0.792
Óbito intraoperatório 7	0.07	0.814
Parada cardíaca prévia/ Taquicardia ventricular instável	0.01	0.921
Tipo de Reoperação 7	0.71	0.928
Tipo de Reoperação 8	0.67	> 0.999
Tipo de Reoperação 9	0.02	> 0.999
Tipo de Dispositivo ao final do procedimento 9	0.95	> 0.999
Óbito intraoperatório 9	0.02	> 0.999
Tipo de Reoperação 10	0.21	> 0.999
Tipo de Dispositivo ao final do procedimento 10	0.21	> 0.999
Óbito intraoperatório 10	0.21	> 0.999
Mudança do tipo de DCEI: entre o Procedimento 8 e Procedimento 9	0.02	> 0.999
Mudança do tipo de DCEI: entre o Procedimento 9 e Procedimento 10	0.21	> 0.999
Readmissão cirúrgica entre 61 a 180 dias	0.00	> 0.999