

# 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
Readmissão em até 30 dias	0.0	< 0.001
Readmissão entre 31 a 60 dias	91354.5	< 0.001
Readmissão entre 61 a 180 dias	267129.0	< 0.001
Readmissão em até 1 ano	424813.5	< 0.001
Número de procedimentos em até 30 dias	3779335.0	< 0.001
Núm. de hospitalizações pós-procedimento	1616924.0	< 0.001
Número de atendimentos	1892867.0	< 0.001
Número de procedimentos em até 60 dias	4288157.0	< 0.001
Óbito durante algum episódio de readmissão hospitalar	3560367.5	< 0.001
Tempo entre o P1 e P2	919330.0	< 0.001
Óbito	3770448.5	< 0.001
Quantidade de classes medicamentosas utilizadas	1564733.5	< 0.001
Quantidade de exames diagnóstico por imagem	2487195.0	< 0.001
Número da Admissão T0	3861791.5	< 0.001
Radiografias	2566380.0	< 0.001
Quantidade de exames por métodos gráficos	2569498.5	< 0.001
ECG	2573005.5	< 0.001
Quantidade de medicamentos de ação cardiovascular	2275320.0	< 0.001
Quantidade de classes medicamentosas de ação cardiovascular	1347954.5	< 0.001
Ecocardiograma	2713427.0	< 0.001
Tempo de seguimento total	5578365.0	< 0.001
Equipe Multiprofissional	2688547.5	< 0.001
Tempo de sobrevida	243918.0	< 0.001

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

Variable	Statistic	p-value
Antiarrítmicos	2550803.0	< 0.001
Quantidade de exames de análises clínicas	2668665.0	< 0.001
Exames laboratoriais	2668808.5	< 0.001
DVA	2517982.0	< 0.001
Diurético	2433199.0	< 0.001
Vasodilatador	2502238.0	< 0.001
Culturas	2956507.0	< 0.001
Núm. de hospitalizações pré-procedimento	4042588.5	< 0.001
Anticoagulantes orais	2757078.0	< 0.001
Antagonista da Aldosterona	2594073.5	< 0.001
Ultrassom	3045991.5	< 0.001
Psicofármacos	2500656.0	< 0.001
Quantidade de procedimentos invasivos	2963279.5	< 0.001
Ano do procedimento 2	675172.0	< 0.001
Quantidade de antimicrobianos	2529161.0	< 0.001
Antibióticos	2533901.0	< 0.001
Antifúngicos	2841473.5	< 0.001
Biopsias	3262107.0	< 0.001
Suporte cardiocirculatório	3275078.5	< 0.001
Cateterismo	3078356.5	< 0.001
Ressonância magnética	3128876.5	< 0.001
Betabloqueador	2741471.5	< 0.001
Insuficiência cardíaca	2657678.5	< 0.001
Cintilografia	3176540.0	< 0.001
Digoxina	2802468.0	< 0.001
Número de comorbidades	4134389.5	< 0.001
Tempo entre o P2 e P3	91723.0	< 0.001
Número de procedimentos em até 180 dias	4555912.5	< 0.001
Cateter venoso central	3209625.0	< 0.001
Tomografia	3109247.5	< 0.001
Número de procedimentos em até 1 ano	4565357.5	< 0.001
Quantidade de exames histopatológicos	3249212.5	< 0.001
Bloqueador do canal de cálcio	2876961.0	< 0.001
Bomba de infusão contínua	2814482.5	< 0.001
Holter	3132523.5	< 0.001
Diárias no serviço de Emergência na admissão T0	1761464.0	< 0.001
Exames endoscópicos	3250420.0	< 0.001
Ano do procedimento 3	87991.5	< 0.001
Antiviral	2923988.0	< 0.001
Estatinas	2715079.5	< 0.001
Transplante cardíaco	3297203.5	< 0.001
Eletrofisiologia	3224269.0	< 0.001
Angio RM	3293722.0	0.001
Óbito hospitalar	4728424.5	0.001
IECA/BRA	2727645.5	0.002
Anticonvulsivante	2879949.0	0.003
Instalação de CEC	3275334.5	0.01
Citologias	3287292.5	0.01
Outros procedimentos cirúrgicos	3224604.0	0.01
Insulina	2870029.0	0.013
Angio TC	3260806.5	0.018
Marca-passo temporário	2865841.0	0.03
Intervenção coronária percutânea	3288415.5	0.032

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

Variable	Statistic	p-value
PET-CT	3295907.0	0.037
Idade no Procedimento 2	592153.5	0.038
Flebografia	3278210.5	0.04
Antiplaquetario EV	2935137.0	0.05
Transfusão de hemoderivados	3288929.5	0.058
Cardioversão/ Desfibrilação	2880533.5	0.092
Intervenção cardiovascular em laboratório de hemodinâmica	3301034.5	0.094
Tilt Test	3306178.5	0.099
Dieta enteral	2882899.0	0.11
Angioplastia	3309421.5	0.117
Teste de esforço	3298521.0	0.181
Ventilação não invasiva	3328148.0	0.206
Arteriografia	3313440.0	0.212
Antihipertensivo	2925031.5	0.266
Antiretroviral	2951160.5	0.268
Angiografia	3311504.5	0.323
Hipoglicemiante	2981659.0	0.341
Idade no momento do primeiro procedimento	4746091.0	0.386
Idade no Procedimento 1	4746091.0	0.386
Polissonografia	3322677.5	0.391
Drenagem de tórax e punção pericárdica ou pleural	3325955.5	0.408
Trombolítico	2957904.0	0.502
Óbito intraoperatório	4652748.0	0.512
Cirurgia Toracica	3313320.5	0.528
Cirurgia Cardiovascular	3334522.5	0.549
Cavografia	3310893.5	0.58
Ano da admissão T0	4689596.0	0.583
Traqueostomia	3315269.0	0.586
Idade no Procedimento 3	73807.0	0.592
Dieta parenteral	2895450.0	0.594
Número de procedimentos na admissão T0	4637873.0	0.613
Ano do procedimento 1	4699863.0	0.65
Stent	3318249.0	0.84
Espirometria / Ergoespirometria	3319581.0	0.87
Número de Mudanças do tipo de DCEI	546525.0	0.923
Interconsulta médica	3316141.5	0.969
Aortografia	3318134.0	0.982
Antiplaquetario VO	2955198.0	NaN
Hormonio tireoidiano	2955198.0	NaN
Broncodilator	2955198.0	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
Insuficiência cardíaca	38.70	< 0.001
Tipo de Procedimento 1	38.90	< 0.001
Tipo de Reoperação 1	51.06	< 0.001
Tipo de Dispositivo ao final do procedimento 1	49.06	< 0.001
Tipo de Reoperação 2	364.37	< 0.001
Óbito intraoperatório 2	22.27	< 0.001
Tipo de Reoperação 3	63.10	< 0.001
Tipo de Dispositivo ao final do procedimento 3	51.92	< 0.001
Tipo de Reoperação 4	44.82	< 0.001
Tipo de Dispositivo ao final do procedimento 4	40.36	< 0.001
Óbito intraoperatório 4	35.96	< 0.001
Mudança do tipo de DCEI: entre o Procedimento 1 e Procedimento 2	21.90	< 0.001
Mudança do tipo de DCEI: entre o Procedimento 2 e Procedimento 3	48.19	< 0.001
Mudança do tipo de DCEI: entre o Procedimento 3 e Procedimento 4	36.20	< 0.001
UTI durante os episódios de hospitalização	100.75	< 0.001
Admissão em até 180 dias antes da T0	49.31	< 0.001
UTI durante a admissão T0	2711.76	< 0.001
Readmissões pós-T0 com diárias de UTI	146.54	< 0.001
Readmissão cirúrgica em até 30 dias	3533.82	< 0.001
Desfecho final do estudo	144.10	< 0.001
Causa do óbito	166.12	< 0.001
Ventilação mecânica / IOT	36.57	< 0.001
Tipo de Dispositivo ao final do procedimento 2	31.35	< 0.001
Óbito intraoperatório 3	48.51	< 0.001
Readmissão cirúrgica em até 1 ano	21.23	< 0.001
Número de procedimentos	62.98	0.001
Desfecho principal da admissão T0	10.25	0.002
Escolaridade	18.94	0.004
Readmissão cirúrgica entre 31 a 60 dias	12.62	0.004
Doença cardíaca	11.67	0.011
Infarto do miocárdio prévio / Doença arterial coronariana	6.95	0.014
Readmissões pós-T0 com diálise	26.40	0.024
Diálise durante os episódios de hospitalização	6.24	0.028
Óbito intraoperatório 5	4.85	0.031
Doença cardíaca	18.60	0.035
Endocardite prévia	4.66	0.043
Classe funcional de IC	11.73	0.045
Hemodiálise	5.93	0.05

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

Variable	Statistic	p-value
Tipo de Reoperação 5	7.50	0.064
Mudança do tipo de DCEI: entre o Procedimento 4 e Procedimento 5	6.15	0.068
Tipo de Dispositivo ao final do procedimento 5	8.57	0.072
Transplante cardíaco prévio	4.85	0.077
Valvopatias/ Prótese valvares	2.49	0.112
Fibrilação / flutter atrial	2.30	0.124
Acidente Vascular Cerebral/ Acidente isquêmico transitório prévios	2.73	0.127
Parada cardíaca prévia/ Taquicardia ventricular instável	1.89	0.169
Insuficiência renal crônica	1.89	0.173
Readmissão cirúrgica entre 61 a 180 dias	2.15	0.189
Diálise durante a admissão T0	33.99	0.191
Neoplasia em tratamento ou tratada recentemente	1.73	0.208
Estado de residência	33.62	0.228
Tipo de Reoperação 6	3.33	0.271
Diabetes melittus	1.25	0.289
Óbito intraoperatório 6	1.30	0.362
Tipo de Dispositivo ao final do procedimento 6	3.41	0.401
Mudança do tipo de DCEI: entre o Procedimento 5 e Procedimento 6	2.11	0.422
Hipertensão arterial	0.58	0.467
Sexo	0.48	0.515
Tipo de Dispositivo ao final do procedimento 7	1.26	0.678
Raça	1.48	0.933
Doença pulmonar obstrutiva crônica	0.01	> 0.999
Óbito intraoperatório 1	0.27	> 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