

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
Número de atendimentos	3091514.5	< 0.001
Núm. de hospitalizações pós-procedimento	3013834.0	< 0.001
Óbito durante algum episódio de readmissão hospitalar	0.0	< 0.001
Óbito em até 2 anos após a alta T0	5536639.0	< 0.001
Óbito em até 3 anos após a alta T0	4499471.5	< 0.001
Óbito	811057.5	< 0.001
Readmissão em até 1 ano	6263772.0	< 0.001
Óbito em até 180 dias após a alta T0	8018939.0	< 0.001
Readmissão entre 61 a 180 dias	7116973.0	< 0.001
Óbito em até 1 ano após a alta T0	7085086.5	< 0.001
Readmissão entre 31 a 60 dias	8104777.0	< 0.001
Readmissão em até 30 dias	8515018.5	< 0.001
Número da Admissão T0	7220850.5	< 0.001
Núm. de hospitalizações pré-procedimento	7415912.0	< 0.001
Tempo de sobrevida	270753.0	< 0.001
Ano da admissão T0	12035820.5	< 0.001
Ano do procedimento 1	12050938.5	< 0.001
Óbito em até 30 dias após a alta T0	9143500.5	< 0.001
Insuficiência cardíaca	8301048.5	< 0.001
Antagonista da Aldosterona	8409636.0	< 0.001
Tempo entre o P1 e P2	1121543.0	< 0.001
Tempo de seguimento total	11258920.0	< 0.001
Ano do procedimento 2	1105326.5	< 0.001
Digoxina	9045142.5	< 0.001
Antiarrítmicos	8653416.5	< 0.001
Número de Mudanças do tipo de DCEI	714846.5	< 0.001
Anticoagulantes orais	9140066.5	< 0.001
Diuretico	8538271.0	< 0.001
Ultrassom	9050489.0	< 0.001
Bomba de infusão contínua	9365042.0	< 0.001
Tempo entre o P2 e P3	128914.5	< 0.001
DVA	8883012.5	< 0.001
Cintilografia	9328048.0	< 0.001
Ano do procedimento 3	122405.5	< 0.001
Holter	9187009.0	< 0.001
Ressonancia magnetica	9283837.5	< 0.001

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

Variable	Statistic	p-value
Interconsulta médica	10031349.5	< 0.001
Equipe Multiprofissional	8949367.5	< 0.001
Ecocardiograma	9021560.5	< 0.001
Antibióticos	10238566.0	< 0.001
Óbito hospitalar	9748909.0	< 0.001
Idade no momento do primeiro procedimento	10201153.0	< 0.001
Idade no Procedimento 1	10201153.0	< 0.001
Eletrofisiologia	9438763.5	< 0.001
Dieta enteral	9566755.0	< 0.001
Dieta parenteral	9590541.0	< 0.001
Antihipertensivo	9774533.5	< 0.001
Número de comorbidades	9114827.5	0.001
Cirurgia Cardiovascular	9759971.5	0.001
Citologias	9546524.0	0.002
Culturas	9336775.0	0.002
Angioplastia	9581678.5	0.002
Tilt Test	9574135.5	0.003
Idade no Procedimento 2	914415.5	0.005
Quantidade de procedimentos invasivos	9333125.0	0.005
Antiplaquetario EV	9560842.0	0.008
Quantidade de exames diagnóstico por imagem	9210190.5	0.011
Cardioversão/ Desfibrilação	9570472.5	0.011
Tomografia	9414966.5	0.011
Vasodilator	9321578.5	0.015
Cateterismo	9430129.5	0.021
Cavografia	9659237.0	0.024
Marca-passo temporário	9557257.5	0.025
Cateter venoso central	9514349.5	0.033
Drenagem de tórax e punção pericárdica ou pleural	9575643.5	0.039
IECA/BRA	9346443.5	0.071
Bloqueador do canal de calcio	9551126.5	0.074
Antifúngicos	9544118.0	0.082
Quantidade de classes medicamentosas utilizadas	9337983.0	0.084
Betabloqueador	9740268.0	0.088
Diárias no serviço de Emergência na admissão T0	9495782.5	0.093
Espirometria / Ergoespirometria	9583088.5	0.103
Quantidade de exames histopatológicos	9566253.5	0.106
Psicofármacos	9392970.5	0.106
Quantidade de exames por métodos gráficos	9363265.0	0.118
Intervenção coronária percutânea	9573331.0	0.119
Radiografias	9378510.5	0.135
Angio TC	9551532.0	0.158
Antiplaquetario VO	9672061.0	0.16
Exames laboratoriais	9397635.5	0.168
Quantidade de exames de análises clínicas	9399702.0	0.172
ECG	9397704.5	0.178
Polissonografia	9597720.0	0.186
Ventilação não invasiva	9627457.0	0.202
Anticonvulsivante	9665106.0	0.239
Estatinas	9462245.0	0.244
Teste de esforço	9638098.0	0.258
PET-CT	9590892.5	0.292
Flebografia	9575847.5	0.301

Table 2: Mann-Whitney Test (continued)

Variable	Statistic	p-value
Trombolítico	9617197.5	0.323
Exames endoscópicos	9581599.5	0.328
Intervenção cardiovascular em laboratório de hemodinâmica	9594426.5	0.365
Antiretroviral	9615892.5	0.372
Hormonio tireoideo	9583795.5	0.402
Broncodilatador	9579913.0	0.432
Outros procedimentos cirúrgicos	9561959.5	0.443
Suporte cardiocirculatório	9599246.5	0.463
Hipoglicemiante	9647022.5	0.464
Instalação de CEC	9628680.0	0.519
Transfusão de hemoderivados	9593026.0	0.519
Angiografia	9602937.0	0.526
Angio RM	9616171.0	0.63
Arteriografia	9607224.0	0.652
Cirurgia Toracica	9615513.0	0.662
Stent	9610672.5	0.766
Transplante cardíaco	9613562.0	0.768
Antiviral	9605884.0	0.799
Insulina	9594522.0	0.804
Traqueostomia	9611785.0	0.835
Idade no Procedimento 3	93415.0	0.838
Aortografia	9611607.5	0.889
Biopsias	9610216.5	0.992

```

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	24.09	< 0.001
Raça	44.23	< 0.001
Escolaridade	181.06	< 0.001
Doença cardíaca	61.67	< 0.001
Classe funcional de IC	140.05	< 0.001
Hipertensão arterial	54.48	< 0.001
Insuficiência cardíaca	87.52	< 0.001
Tipo de Procedimento 1	17.72	< 0.001
Tipo de Dispositivo ao final do procedimento 1	259.35	< 0.001
Tipo de Reoperação 2	64.19	< 0.001
Tipo de Dispositivo ao final do procedimento 2	165.99	< 0.001
Tipo de Reoperação 3	45.74	< 0.001
Tipo de Dispositivo ao final do procedimento 3	63.92	< 0.001
Óbito intraoperatório 3	48.11	< 0.001
Mudança do tipo de DCEI: entre o Procedimento 1 e Procedimento 2	66.71	< 0.001
Díalise durante os episódios de hospitalização	68.08	< 0.001
UTI durante os episódios de hospitalização	213.95	< 0.001
Admissão em até 180 dias antes da T0	170.45	< 0.001
UTI durante a admissão T0	2323.39	< 0.001
Readmissões pós-T0 com diárias de UTI	317.85	< 0.001
Readmissões pós-T0 com diálise	125.31	< 0.001
Desfecho principal da admissão T0	23.23	< 0.001
Desfecho final do estudo	6716.91	< 0.001
Causa do óbito	749.91	< 0.001
Ventilação mecânica / IOT	18.05	< 0.001
Doença cardíaca	20.44	< 0.001
Insuficiência renal crônica	12.04	< 0.001
Tipo de Reoperação 1	20.84	< 0.001
Mudança do tipo de DCEI: entre o Procedimento 2 e Procedimento 3	16.65	0.002
Tipo de Reoperação 4	15.78	0.003
Tipo de Dispositivo ao final do procedimento 4	18.18	0.004
Parada cardíaca prévia/ Taquicardia ventricular instável	8.86	0.005
Fibrilação / flutter atrial	7.61	0.006
Valvopatias/ Prótese valvares	7.25	0.006
Endocardite prévia	7.52	0.006
Tipo de Dispositivo ao final do procedimento 6	17.41	0.007
Mudança do tipo de DCEI: entre o Procedimento 5 e Procedimento 6	12.80	0.008
Óbito intraoperatório	11.73	0.009
Estado de residência	54.35	0.013
Transplante cardíaco prévio	6.32	0.019
Óbito intraoperatório 4	5.20	0.024
Díalise durante a admissão T0	48.42	0.026
Mudança do tipo de DCEI: entre o Procedimento 3 e Procedimento 4	7.11	0.029
Mudança do tipo de DCEI: entre o Procedimento 4 e Procedimento 5	8.01	0.034
Óbito intraoperatório 6	4.65	0.043
Infarto do miocárdio prévio / Doença arterial coronariana	4.16	0.052
Tipo de Reoperação 6	8.24	0.052
Número de procedimentos	19.35	0.052
Mudança do tipo de DCEI: entre o Procedimento 7 e Procedimento 8	11.30	0.092
Tipo de Dispositivo ao final do procedimento 7	11.65	0.092
Óbito intraoperatório 2	2.31	0.134
Doença pulmonar obstrutiva crônica	2.10	0.171
Diabetes melittus	1.64	0.206

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

Variable	Statistic	p-value
Tipo de Dispositivo ao final do procedimento 5	5.50	0.227
Tipo de Dispositivo ao final do procedimento 8	1.17	0.248
Acidente Vascular Cerebral/ Acidente isquêmico transitório prévios	1.47	0.249
Tipo de Reoperação 8	1.17	0.249
Óbito intraoperatório 8	1.17	0.267
Óbito intraoperatório 5	1.35	0.294
Mudança do tipo de DCEI: entre o Procedimento 8 e Procedimento 9	0.94	0.336
Mudança do tipo de DCEI: entre o Procedimento 6 e Procedimento 7	2.58	0.343
Tipo de Reoperação 9	0.94	0.347
Tipo de Dispositivo ao final do procedimento 9	0.94	0.349
Óbito intraoperatório 9	0.94	0.364
Hemodiálise	0.89	0.404
Óbito intraoperatório 7	0.81	0.498
Tipo de Reoperação 7	1.78	0.61
Tipo de Reoperação 5	1.75	0.614
Óbito intraoperatório 1	0.62	0.653
Neoplasia em tratamento ou tratada recentemente	0.04	0.856
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