script_trab2.R

p

2022-07-08

```
rm(list=ls())
dev.off()
```

```
## null device
##           1
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.7     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(magrittr)
```

```
##
## Attaching package: 'magrittr'
```

```
## The following object is masked from 'package:purrr':
##
##     set_names
```

```
## The following object is masked from 'package:tidyr':
##
##     extract
```

```
addu <- adu <- read.csv("~/Documents/UFScar 202201/Modelos de Regressão/Trabalho 02/data/adult.data",
                  header=FALSE, stringsAsFactors=FALSE) %>% mutate_if(is.character, str_trim)

limit <- 500000

"Prediction task is to determine whether a person makes over 50K a year.

age: continuous.
workclass: Private, Self-emp-not-inc, Self-emp-inc, Federal-gov, Local-gov, State-gov,
    Without-pay, Never-worked.
fnlwgt: continuous.
education: Bachelors, Some-college, 11th, HS-grad, Prof-school, Assoc-acdm, Assoc-voc,
    9th, 7th-8th, 12th, Masters, 1st-4th, 10th, Doctorate, 5th-6th, Preschool.
education-num: continuous.
marital-status: Married-civ-spouse, Divorced, Never-married, Separated, Widowed,
```

```
    Married-spouse-absent, Married-AF-spouse.
occupation: Tech-support, Craft-repair, Other-service, Sales, Exec-managerial, Prof-specialty,
    Handlers-cleaners, Machine-op-inspct, Adm-clerical, Farming-fishing, Transport-moving,
    Priv-house-serv, Protective-serv, Armed-Forces.
relationship: Wife, Own-child, Husband, Not-in-family, Other-relative, Unmarried.
race: White, Asian-Pac-Islander, Amer-Indian-Eskimo, Other, Black.
sex: Female, Male.
capital-gain: continuous.
capital-loss: continuous.
hours-per-week: continuous.
native-country: United-States, Cambodia, England, Puerto-Rico, Canada, Germany,
    Outlying-US(Guam-USVI-etc), India, Japan, Greece, South, China, Cuba, Iran,
    Honduras, Philippines, Italy, Poland, Jamaica, Vietnam, Mexico, Portugal,
    Ireland, France, Dominican-Republic, Laos, Ecuador, Taiwan, Haiti, Columbia,
    Hungary, Guatemala, Nicaragua, Scotland, Thailand, Yugoslavia, El-Salvador,
    Trinadad&Tobago, Peru, Hong, Holand-Netherlands."
```

```
## [1] "Prediction task is to determine whether a person makes over 50K a year.\n\nage: continuous.\nwo
```

```
columns <- c("age", "work", "fnlwgt", "education", "education_num", "marital_status",
             "occupation", "relationship", "race", "sex", "capital_gain", "capital_loss",
             "hours_per_week", "native_country", "group")
colunas <- c("idade", "trabalho", "renda_anual", "educacao", "anos_educacao", "estado_civil",
             "profissicao", "nao_sei", "race", "sex", "capital_ganho", "capital_perda",
             "horas_por_semana", "nacionalidade", "y")
names(adu) <- abbreviate(colunas)
names(adu)
```

```
##  [1] "idad"   "trbl"   "rnd_"   "edcc"   "ans_"   "est_"   "prfs"   "na_s"
##  [9] "race"   "sex"    "cptl_g" "cptl_p" "hr__"   "ncnl"   "y"
```

```
addu <- adu
adu <- addu %>% sample_n(100)


# Agrupamento de variaveis

quais <- (
  lapply(
    lapply(select_if(addu, is.character),
           table), length) > 5) %>%
  which %>% names
func <- paste0("unique(addu$", quais ,")")

eval(parse(text = func[1]))
```

```
## [1] "State-gov"        "Self-emp-not-inc" "Private"          "Federal-gov"
## [5] "Local-gov"        "?"                "Self-emp-inc"     "Without-pay"
## [9] "Never-worked"
```

```
lista <- func %>%
  map(function(x) parse(text = x)) %>%
  map(.f = eval, .x = .)
names(lista) <- quais

quais[1]
```

```
## [1] "trbl"
```

```
# Local de trabalho ####

addu %>% group_by(trbl) %>% summarise(n = 100*n()/nrow(addu)) %>% arrange(n)
```

```
## # A tibble: 9 x 2
##   trbl                 n
##   <chr>            <dbl>
## 1 Never-worked    0.0215
## 2 Without-pay     0.0430
## 3 Federal-gov     2.95
## 4 Self-emp-inc    3.43
## 5 State-gov       3.99
## 6 ?               5.64
## 7 Local-gov       6.43
## 8 Self-emp-not-inc  7.80
## 9 Private        69.7
```

```
ggplot(addu, aes(x = 1:nrow(addu), y = rnd_, col = trbl)) +
  geom_point() +
  geom_hline(yintercept = limit) +
  facet_wrap(~trbl) +
  theme_bw()

prop.table(table(addu$rnd_<limit, addu$trbl), margin = 2)
```

```
##
##                    ? Federal-gov   Local-gov Never-worked     Private
##   FALSE 0.011437908 0.013541667 0.010033445  0.000000000 0.012381036
##   TRUE  0.988562092 0.986458333 0.989966555  1.000000000 0.987618964
##
##         Self-emp-inc Self-emp-not-inc   State-gov Without-pay
##   FALSE  0.004480287      0.005509642 0.013097072 0.000000000
##   TRUE   0.995519713      0.994490358 0.986902928 1.000000000
```

```
addu$trbl %>% unique
```

```
## [1] "State-gov"        "Self-emp-not-inc" "Private"          "Federal-gov"
## [5] "Local-gov"        "?"                "Self-emp-inc"     "Without-pay"
## [9] "Never-worked"
```

```
addu <- addu %>% mutate(trbl = ifelse(trbl %in% c("Never-worked", "Without-pay", "?"),
                                      "Others", trbl)) %>% # Juntando pequena variaveis
  mutate(trbl = ifelse(str_detect(trbl, "gov"), "Gov", trbl)) %>% # Juntando func publ
  mutate(trbl = ifelse(str_detect(trbl, "emp"), "Self-emp", trbl)) # Juntando empresarios
ggplot(addu, aes(x = 1:nrow(addu), y = rnd_, col = trbl)) +
  geom_point() +
  geom_hline(yintercept = limit) +
  facet_wrap(~trbl) +
  theme_bw()

quais[2]
```

```
## [1] "edcc"
```

```
# Transformando Variavel educacao ####


addu %>% group_by(edcc) %>% summarise(n = 100*n()/nrow(addu)) %>% arrange(n)
```

```
## # A tibble: 16 x 2
##    edcc              n
##    <chr>         <dbl>
##  1 Preschool     0.157
##  2 1st-4th       0.516
##  3 5th-6th       1.02
##  4 Doctorate     1.27
##  5 12th          1.33
##  6 9th           1.58
##  7 Prof-school   1.77
##  8 7th-8th       1.98
##  9 10th          2.87
## 10 Assoc-acdm    3.28
## 11 11th          3.61
## 12 Assoc-voc     4.24
## 13 Masters       5.29
## 14 Bachelors    16.4
## 15 Some-college 22.4
## 16 HS-grad      32.3
```

```
ggplot(addu, aes(x = 1:nrow(addu), y = rnd_, col = edcc)) +
  geom_point() +
  geom_hline(yintercept = limit) +
  facet_wrap(~edcc) +
  theme_bw()

table(addu$rnd_<limit, addu$edcc) %>% prop.table(margin = 2)
```

```
##
##                10th        11th        12th     1st-4th     5th-6th     7th-8th
##   FALSE 0.017148982 0.010212766 0.025404157 0.029761905 0.030030030 0.007739938
##   TRUE  0.982851018 0.989787234 0.974595843 0.970238095 0.969969970 0.992260062
##
##                 9th  Assoc-acdm   Assoc-voc   Bachelors   Doctorate     HS-grad
##   FALSE 0.017509728 0.004686036 0.005788712 0.012324930 0.007263923 0.011522712
##   TRUE  0.982490272 0.995313964 0.994211288 0.987675070 0.992736077 0.988477288
##
##             Masters   Preschool Prof-school Some-college
##   FALSE 0.010446895 0.039215686 0.003472222  0.010835276
##   TRUE  0.989553105 0.960784314 0.996527778  0.989164724
```

```
plot(table(addu$edcc, addu$rnd_<limit), las = 2)
addu$edcc %>% unique %>% cat(sep = "', '")
```

```
## Bachelors', 'HS-grad', '11th', 'Masters', '9th', 'Some-college', 'Assoc-acdm', 'Assoc-voc', '7th-8th
```

```
fund <- c('Preschool', '1st-4th', '5th-6th', '7th-8th', '9th',  '10th', '11th')
lib <- c('Assoc-acdm', 'Prof-school')
grad <- c('12th', 'HS-grad', 'Masters', 'Doctorate', 'Assoc-voc', 'Some-college', 'Bachelors')

addu <- addu %>%
```

```
  mutate(edcc = ifelse(edcc %in% fund, "Fundamental", edcc)) %>%
  mutate(edcc = ifelse(edcc %in% grad, "Grad", edcc)) %>%
  mutate(edcc = ifelse(edcc %in% lib, "Liberal", edcc))

ggplot(addu, aes(x = 1:nrow(addu), y = rnd_, col = edcc)) +
  geom_point() +
  geom_hline(yintercept = limit) +
  facet_wrap(~edcc) +
  theme_bw()



quais[3]
```

```
## [1] "est_"
```

```
which(names(adu)==quais[3])
```

```
## [1] 6
```

```
# Transformando Variavel estado civil ####


addu %>% group_by(est_) %>% summarise(n = 100*n()/nrow(addu)) %>% arrange(n)
```

```
## # A tibble: 7 x 2
##   est_                      n
##   <chr>                 <dbl>
## 1 Married-AF-spouse      0.0706
## 2 Married-spouse-absent  1.28
## 3 Widowed                3.05
## 4 Separated              3.15
## 5 Divorced              13.6
## 6 Never-married         32.8
## 7 Married-civ-spouse    46.0
```

```
ggplot(addu, aes(x = 1:nrow(addu), y = rnd_, col = est_)) +
  geom_hline(yintercept = limit) +
  geom_point() + facet_wrap(~est_) +
  theme_bw()

table(addu$rnd_<limit, addu$est_) %>% prop.table(margin = 2)
```

```
##
##           Divorced Married-AF-spouse Married-civ-spouse Married-spouse-absent
##   FALSE 0.010353365       0.043478261        0.010483440           0.014354067
##   TRUE  0.989646635       0.956521739        0.989516560           0.985645933
##
##         Never-married   Separated     Widowed
##   FALSE   0.012730506 0.018536585 0.007049345
##   TRUE    0.987269494 0.981463415 0.992950655
```

```
plot(table(addu$est_, addu$rnd_<limit), las = 2)
addu$est_ %>% unique %>% cat(sep = "', '")
```

```
## Never-married', 'Married-civ-spouse', 'Divorced', 'Married-spouse-absent', 'Separated', 'Married-AF-s
```

```
fund <- c('Preschool', '1st-4th', '5th-6th', '7th-8th', '9th',  '10th', '11th')
lib <- c('Assoc-acdm', 'Prof-school')
grad <- c('12th', 'HS-grad', 'Masters', 'Doctorate', 'Assoc-voc', 'Some-college', 'Bachelors')

marrd <- c('Married-civ-spouse', 'Divorced', 'Married-spouse-absent', 'Married-AF-spouse', 'Widowed')
addu <- addu %>%
  mutate(est_ = ifelse(est_ %in% marrd, "Married", est_))

ggplot(addu, aes(x = 1:nrow(addu), y = rnd_, col = est_)) +
  geom_point() +
  geom_hline(yintercept = limit) +
  facet_wrap(~est_) +
  theme_bw()

quais[4]
```

```
## [1] "prfs"
```

```
which(names(adu)==quais[4])
```

```
## [1] 7
```

```
# Transformando variavel profissao ####


addu %>% group_by(prfs) %>% summarise(n = 100*n()/nrow(addu)) %>% arrange(n)
```

```
## # A tibble: 15 x 2
##    prfs                  n
##    <chr>             <dbl>
##  1 Armed-Forces     0.0276
##  2 Priv-house-serv  0.458
##  3 Protective-serv  1.99
##  4 Tech-support     2.85
##  5 Farming-fishing  3.05
##  6 Handlers-cleaners  4.21
##  7 Transport-moving  4.90
##  8 ?                5.66
##  9 Machine-op-inspct  6.15
## 10 Other-service    10.1
## 11 Sales            11.2
## 12 Adm-clerical     11.6
## 13 Exec-managerial  12.5
## 14 Craft-repair     12.6
## 15 Prof-specialty   12.7
```

```
ggplot(addu, aes(x = 1:nrow(addu), y = rnd_, col = prfs)) +
  geom_hline(yintercept = limit) +
  geom_point() + facet_wrap(~prfs) +
  theme_bw()

table(addu$rnd_<limit, addu$prfs) %>% prop.table(margin = 2) %>%
  as.data.frame.matrix() %>% t %>%
  data.frame() %>% rownames_to_column(var = "prfs") %>%
  inner_join(.,
             addu %>% group_by(prfs) %>% summarise(n = 100*n()/nrow(addu)) %>% arrange(n)) %>%
```

```
  arrange(desc(FALSE.))
```

```
## Joining, by = "prfs"
```

```
##                  prfs      FALSE.     TRUE.           n
## 1  Handlers-cleaners 0.017518248 0.9824818  4.20748749
## 2   Transport-moving 0.015028178 0.9849718  4.90464052
## 3    Farming-fishing 0.014084507 0.9859155  3.05273180
## 4       Tech-support 0.014008621 0.9859914  2.85003532
## 5    Priv-house-serv 0.013422819 0.9865772  0.45760265
## 6       Craft-repair 0.012929983 0.9870700 12.58867971
## 7       Adm-clerical 0.012732095 0.9872679 11.57826848
## 8    Protective-serv 0.012326656 0.9876733  1.99318203
## 9      Other-service 0.011532625 0.9884674 10.11946808
## 10                 ? 0.011394466 0.9886055  5.66014557
## 11 Machine-op-inspct 0.010989011 0.9890110  6.14845981
## 12   Exec-managerial 0.009591736 0.9904083 12.48733147
## 13             Sales 0.009589041 0.9904110 11.20972943
## 14    Prof-specialty 0.007487923 0.9925121 12.71459722
## 15      Armed-Forces 0.000000000 1.0000000  0.02764043
```

```
plot(table(addu$prfs, addu$rnd_<limit), las = 2)
addu$prfs %>% unique %>% cat(sep = "', '")
```

```
## Adm-clerical', 'Exec-managerial', 'Handlers-cleaners', 'Prof-specialty', 'Other-service', 'Sales', '
```

```
produ <- c('Handlers-cleaners', 'Transport-moving', 'Farming-fishing', 'Tech-support')
admin <- c('Exec-managerial', 'Adm-clerical')
servi <- c('Priv-house-serv', 'Craft-repair', 'Protective-serv', 'Machine-op-inspct')

addu <- addu %>%
  mutate(classe = 'Others') %>%
  mutate(classe = ifelse(prfs %in% produ, "Production", classe)) %>%
  mutate(classe = ifelse(prfs %in% admin, "Administrative", classe)) %>%
  mutate(classe = ifelse(prfs %in% servi, "Service", classe))

ggplot(addu, aes(x = 1:nrow(addu), y = rnd_, col = classe)) +
  geom_point() + facet_wrap(~classe) +
  theme_bw()

addu$prfs <- NULL

quais[5]
```

```
## [1] "na_s"
```

```
which(names(adu)==quais[5])
```

```
## [1] 8
```

```
# Retirando variavel relacionamentos ####
# nao consigo ver como essas categorias sao mutuamente excludentes

addu$na_s <- NULL


quais[6]
```

```
## [1] "ncnl"
```

```
which(names(adu)==quais[6])
```

```
## [1] 14
```

```
# Transformando variavel paises em continentes ####


addu %>% group_by(ncnl) %>% summarise(n = 100*n()/nrow(addu)) %>% arrange(n)
```

```
## # A tibble: 42 x 2
##    ncnl                          n
##    <chr>                       <dbl>
##  1 Holand-Netherlands        0.00307
##  2 Scotland                  0.0369
##  3 Honduras                  0.0399
##  4 Hungary                   0.0399
##  5 Outlying-US(Guam-USVI-etc) 0.0430
##  6 Yugoslavia                0.0491
##  7 Laos                      0.0553
##  8 Thailand                  0.0553
##  9 Cambodia                  0.0584
## 10 Trinadad&Tobago           0.0584
## # ... with 32 more rows
```

```
ggplot(addu, aes(x = 1:nrow(addu), y = rnd_, col = ncnl)) +
  geom_hline(yintercept = limit) +
  geom_point() + facet_wrap(~ncnl) +
  theme_bw()

table(addu$rnd_<limit, addu$ncnl) %>% prop.table(margin = 2) %>%
  as.data.frame.matrix() %>% t %>%
  data.frame() %>% rownames_to_column(var = "ncnl") %>%
  inner_join(.,
             addu %>% group_by(ncnl) %>% summarise(n = 100*n()/nrow(addu)) %>% arrange(n)) %>%
  arrange(desc(FALSE.))
```

```
## Joining, by = "ncnl"
```

```
##                          ncnl      FALSE.     TRUE.            n
## 1                   Nicaragua 0.117647059 0.8823529  0.104419397
## 2                        Peru 0.096774194 0.9032258  0.095205921
## 3                 El-Salvador 0.066037736 0.9339623  0.325542827
## 4                   Guatemala 0.062500000 0.9375000  0.196554160
## 5                  Yugoslavia 0.062500000 0.9375000  0.049138540
## 6              Trinadad&Tobago 0.052631579 0.9473684  0.058352016
## 7                      Mexico 0.048211509 0.9517885  1.974755075
## 8                    Columbia 0.033898305 0.9661017  0.181198366
## 9                       Japan 0.032258065 0.9677419  0.190411842
## 10                     Taiwan 0.019607843 0.9803922  0.156629096
## 11                Puerto-Rico 0.017543860 0.9824561  0.350112097
## 12                          ? 0.015437393 0.9845626  1.790485550
## 13                      Italy 0.013698630 0.9863014  0.224194589
## 14                       Cuba 0.010526316 0.9894737  0.291760081
## 15              United-States 0.010284539 0.9897155 89.585700685
## 16                     Canada 0.008264463 0.9917355  0.371610209
```

```
## 17                    Germany 0.007299270 0.9927007  0.420748749
## 18                Philippines 0.005050505 0.9949495  0.608089432
## 19                   Cambodia 0.000000000 1.0000000  0.058352016
## 20                      China 0.000000000 1.0000000  0.230336906
## 21         Dominican-Republic 0.000000000 1.0000000  0.214981112
## 22                    Ecuador 0.000000000 1.0000000  0.085992445
## 23                    England 0.000000000 1.0000000  0.276404287
## 24                     France 0.000000000 1.0000000  0.089063604
## 25                     Greece 0.000000000 1.0000000  0.089063604
## 26                      Haiti 0.000000000 1.0000000  0.135130985
## 27         Holand-Netherlands 0.000000000 1.0000000  0.003071159
## 28                   Honduras 0.000000000 1.0000000  0.039925064
## 29                       Hong 0.000000000 1.0000000  0.061423175
## 30                    Hungary 0.000000000 1.0000000  0.039925064
## 31                      India 0.000000000 1.0000000  0.307115875
## 32                       Iran 0.000000000 1.0000000  0.132059826
## 33                    Ireland 0.000000000 1.0000000  0.073707810
## 34                    Jamaica 0.000000000 1.0000000  0.248763859
## 35                       Laos 0.000000000 1.0000000  0.055280857
## 36 Outlying-US(Guam-USVI-etc) 0.000000000 1.0000000  0.042996222
## 37                     Poland 0.000000000 1.0000000  0.184269525
## 38                   Portugal 0.000000000 1.0000000  0.113632874
## 39                   Scotland 0.000000000 1.0000000  0.036853905
## 40                      South 0.000000000 1.0000000  0.245692700
## 41                   Thailand 0.000000000 1.0000000  0.055280857
## 42                    Vietnam 0.000000000 1.0000000  0.205767636
```

```
plot(table(addu$ncnl, addu$rnd_<limit), las = 2)
addu$ncnl %>% unique %>% sort %>% cat(sep = "', '")
```

```
## ?', 'Cambodia', 'Canada', 'China', 'Columbia', 'Cuba', 'Dominican-Republic', 'Ecuador', 'El-Salvador
```

```
asia <- c('Cambodia', 'China', 'Hong', 'India', 'Japan', 'Laos', 'Philippines', 'Taiwan',
          'Thailand', 'Vietnam')
americas <- c('Canada', 'Columbia', 'Cuba', 'Dominican-Republic', 'Ecuador', 'El-Salvador',
              'Guatemala', 'Haiti', 'Honduras', 'Jamaica', 'Mexico', 'Nicaragua', 'Peru',
              'Puerto-Rico', 'United-States')
europe <- c('England', 'France', 'Germany', 'Greece', 'Holand-Netherlands', 'Hungary',
            'Ireland', 'Italy', 'Poland', 'Portugal', 'Scotland', 'Trinadad&Tobago',
            'Yugoslavia')

abbreviate('continent')
```

```
## continent
##    "cntn"
```

```
addu <- addu %>%
  mutate(cntn = 'Others') %>%
  mutate(cntn = ifelse(ncnl %in% asia, "Asia", cntn)) %>%
  mutate(cntn = ifelse(ncnl %in% americas, "America", cntn)) %>%
  mutate(cntn = ifelse(ncnl %in% europe, "Europe", cntn))

ggplot(addu, aes(x = 1:nrow(addu), y = rnd_, col = cntn)) +
  geom_point() +
  facet_wrap(~cntn) +
  theme_bw()
```

```
addu$ncnl <- NULL

# Revisando variaveis categoricas ####

quais1 <- (
  lapply(
    lapply(select_if(addu, is.character),
           table), length) > 5) %>%
  which %>% names

quais1
```

```
## character(0)
```

```
# Decidindo qual salario sera previsto ####

plot(adu$rnd_ %>% sort, pch = 16)

x <- kmeans(adu$rnd_ %>% sort, 4)
plot(adu$rnd_ %>% sort, col = x$cluster, pch = 16)

# Limpar memoria! ####

rm(list = ls()[which(!ls() %in% c("addu", "adu"))])
dev.off()
```

```
## null device
##           1
```

```
# Produzindo modelos encaixados ####

adu <- mutate_if(select(adu, -rnd_), is.character, as.factor)
addu <- mutate_if(select(addu, -rnd_), is.character, as.factor)

mod_sat <- glm(y~., data = adu, family = 'binomial')
```

```
## Warning: glm.fit: algorithm did not converge
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
index <- names(adu)[-which(names(adu)=='y')]

# Cria um modelo
first_var <- function(x, varr){
  formula <- paste0("y~",varr)
  mod1 <- glm(formula, data = x, family = binomial)
}

# Cria os modelos possivel com 1 variavel
mods1 <- map(x = adu, .x = index, .f = first_var)
names(mods1) <- index

# Pega o modelo com o melhor logLik
varr <- lapply(mods1, logLik) %>% unlist %>% sort(decreasing = TRUE) %>% .[1] %>% names

mods1[[varr]]
```

```
##
## Call:  glm(formula = formula, family = binomial, data = x)
##
## Coefficients:
##        (Intercept)   na_sNot-in-family  na_sOther-relative       na_sOwn-child
##            -0.4700             -2.8258            -19.0961            -19.0961
##      na_sUnmarried            na_sWife
##           -19.0961             -0.2231
##
## Degrees of Freedom: 99 Total (i.e. Null);  94 Residual
## Null Deviance:       91.18
## Residual Deviance: 64.42     AIC: 76.42
```

```
#
# mod1 <- glm(formula, data = adu, family = binomial)
#
# 2*(logLik(mod_sat) - logLik(mod))
#
# y <- lapply(X = index, FUN = first_var, x = adu)
# names(y) <- index
#
# library(statmod)
# qresiduals(logit)
# qres.binom(glm.obj)
#
# GGally::ggpairs(adu %>% select(-edcc, -ncnl))
# cor(adu)
# predict <- predict(logit, adu, type = 'response')
# qresid
#
# library(surveillance)
#
# anscombe.residuals(logit, sigma(logit)) %>% hist
# resid(logit) %>% hist
#
#
# # Curva ROC ####
#
# library(ROCR)
# ROCRpred <- prediction(predict, adu$grup)
# ROCRperf <- performance(ROCRpred, 'tpr', 'fpr')
# plot(ROCRperf, colorize = TRUE, text.adj = c(-0.2, 1.7))
# beepr::beep(4)
#
# print.AsIs(ROCRperf)
#
#
#
#
#
# plot(x = unlist(ROCRperf@x.values), y = unlist(ROCRperf@y.values))
#
# data.frame(x = unlist(ROCRperf@x.values), y = unlist(ROCRperf@y.values)) %>%
#   mutate(diss = sqrt(x^2 + (y-1)^2)) %>% arrange((diss)) %>% glimpse
#
```

```
#
# # Confusion matrix ####
#
# table_mat <- table(adu$grup, predict > 0.846)
# prop.table(table_mat)
#
#
# score <- qlogis(logit$fitted.values)
# class <- adu$grup
# library(ROCit)
# rocit_emp <- rocit(score = score,
#                    class = class,
#                    method = "emp")
# rocit_bin <- rocit(score = score,
#                    class = class,
#                    method = "bin")
# rocit_non <- rocit(score = score,
#                    class = class,
#                    method = "non")
# summary(rocit_emp)
# summary(rocit_bin)
# summary(rocit_non)
#
```