

Análise Descritiva de Dados com R : Base de microdados enade 2018

Gilson Castro

27/08/2020

Descrição do problema

O coordenador do curso de ciências econômicas de uma faculdade no nordeste solicitou que sua equipe de trabalho fizesse um estudo sobre o desempenho dos alunos do curso mas para isso, seria preciso entender o conjunto de dados dessa forma, foi escolhida a base de microdados do enade , afim de resumir as notas geral para elaborar estratégias pedagógicas com objetivo de melhorar o desempenho da sua instituição.

Três questões para responder

1. Onde estão concentradas as maiores notas por turno ?
2. Qual turno tem as maiores notas?
3. Qual Estado tem a maior nota média ?

Base de dados

Para realizar a análise descritiva o grupo de trabalho utilizou uma base de [microdados](#)(**dados desagregados**) do [ENADE](#) (**Exame Nacional de Desempenho dos Estudantes**) esse exame avalia o desempenho dos alunos de ensino superior concluintes dos cursos de graduação , através de um questionário composto por perguntas relacionadas ao perfil do aluno , conteúdo programático do curso , habilidade , competências .

Tipo de análise de dados : **Análise descritiva**

Nesta análise o objetivo é organizar e descrever o conjunto de dados , esse tipo de análise é um dos primeiros passos da análise estatística pois auxilia na visão geral dos dados utilizando as medidas de posição e dispersão . Será apresentado também gráficos e tabelas para melhor compreensão do problema.

Carregando os pacotes da análise

```
library(ggplot2)
library(readr)
library(e1071)
library(plotly)
```

```
##
## Attaching package: 'plotly'
```

```
## The following object is masked from 'package:ggplot2':
##
##     last_plot
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

```
## The following object is masked from 'package:graphics':
##
##     layout
```

```
require(dplyr)
```

```
## Loading required package: dplyr
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##   filter, lag
```

```
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
require(Hmisc)
```

```
## Loading required package: Hmisc
```

```
## Loading required package: lattice
```

```
## Loading required package: survival
```

```
## Loading required package: Formula
```

```
##
## Attaching package: 'Hmisc'
```

```
## The following objects are masked from 'package:dplyr':
##
##   src, summarize
```

```
## The following object is masked from 'package:plotly':
##
##   subplot
```

```
## The following object is masked from 'package:e1071':
##
##   impute
```

```
## The following objects are masked from 'package:base':
##
##   format.pval, units
```

```
require(esquisse)
```

```
## Loading required package: esquisse
```

```
require(devtools)
```

```
## Loading required package: devtools
```

```
## Loading required package: usethis
```

```
require(readr)
require(ggplot2)
require(e1071)
require(plotly)
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.0 --
```

```
## v tibble 3.0.3      v stringr 1.4.0
## v tidyr  1.1.0      v forcats 0.5.0
## v purrr  0.3.4
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter()      masks plotly::filter(), stats::filter()
## x dplyr::lag()         masks stats::lag()
## x Hmisc::src()         masks dplyr::src()
## x Hmisc::summarize()   masks dplyr::summarize()
```

```
library(scales)
```

```
##
## Attaching package: 'scales'
```

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

```
## The following object is masked from 'package:readr':
##
##     col_factor
```

Carregando os dados

```
#verificação e alteração de pasta destino para trabalhar
setwd("C:/Users/gilso/OneDrive/Documentos/enade2018")
getwd()
```

```
## [1] "C:/Users/gilso/OneDrive/Documentos/enade2018"
```

```
#carregando a base de dados
base_enade=read.table("microdados_enade_2018.txt",header = TRUE,sep = ";",dec = ",",colClasses = c(NT_OBJ_FG=
"numeric"))
#visualizando completa a base
View(base_enade)
# verificação da estrutura do conjunto de dados
str(base_enade)
```

```
## 'data.frame':    548127 obs. of  137 variables:
## $ NU_ANO          : int  2018 2018 2018 2018 2018 2018 2018 2018 2018 2018 ...
## $ CO_IES          : int  1356 1356 1356 1356 1356 1356 1356 1356 1356 1356 ...
## $ CO_CATEGAD      : int  10003 10003 10003 10003 10003 10003 10003 10003 10003 10003 ...
## $ CO_ORGACAD      : int  10020 10020 10020 10020 10020 10020 10020 10020 10020 10020 ...
## $ CO_GRUPO        : int  1 1 1 1 1 1 1 1 1 1 ...
## $ CO_CURSO        : int  47116 47116 47116 47116 47116 47116 47116 47116 47116 47116 ...
## $ CO_MODALIDADE   : int  1 1 1 1 1 1 1 1 1 1 ...
## $ CO_MUNIC_CURSO  : int  3546603 3546603 3546603 3546603 3546603 3546603 3546603 3546603 3546603 3546603 ...
## $ CO_UF_CURSO     : int  35 35 35 35 35 35 35 35 35 35 ...
## $ CO_REGIAO_CURSO : int  3 3 3 3 3 3 3 3 3 3 ...
## $ NU_IDADE        : int  22 49 23 25 22 22 22 22 25 21 ...
## $ TP_SEXO         : chr  "M" "F" "M" "M" ...
## $ ANO_FIM_EM      : int  2013 1988 2013 2011 2014 2014 2014 2014 2011 2014 ...
## $ ANO_IN_GRAD     : int  2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 ...
## $ CO_TURNO_GRADUACAO : int  4 4 4 4 4 4 4 4 4 4 ...
## $ TP_INSCRICAO_ADM : int  0 0 0 0 0 0 0 0 0 0 ...
## $ TP_INSCRICAO     : int  0 0 0 0 0 0 0 0 0 0 ...
## $ NU_ITEM_OFG      : int  8 8 8 8 8 8 8 8 8 8 ...
## $ NU_ITEM_OFG_Z     : int  0 0 0 0 0 0 0 0 0 0 ...
## $ NU_ITEM_OFG_X     : int  0 0 0 0 0 0 0 0 0 0 ...
## $ NU_ITEM_OFG_N     : int  0 0 0 0 0 0 0 0 0 0 ...
## $ NU_ITEM_OCE      : int  27 27 27 27 27 27 27 27 27 27 ...
## $ NU_ITEM_OCE_Z     : int  1 1 1 1 1 1 1 1 1 1 ...
## $ NU_ITEM_OCE_X     : int  8 8 8 8 8 8 8 8 8 8 ...
## $ NU_ITEM_OCE_N     : int  0 0 0 0 0 0 0 0 0 0 ...
## $ DS_VT_GAB_OFG_ORIG : chr  "CACBEBDE" "CACBEBDE" "CACBEBDE" "CACBEBDE" ...
## $ DS_VT_GAB_OFG_FIN : chr  "CACBEBDE" "CACBEBDE" "CACBEBDE" "CACBEBDE" ...
## $ DS_VT_GAB_OCE_ORIG : chr  "ADBBDEDCEEDCAEZADBCBDABECBB" "ADBBDEDCEEDCAEZADBCBDABECBB" "ADBBDEDCEEDCAEZADBCBDABECBB" "ADBBDEDCEEDCAEZADBCBDABECBB" ...
```

```
## $ DS_VT_GAB_OCE_FIN : chr "XXBXDEDXEEDCAEZADBCBXAXECXX" "XXBXDEDXEEDCAEZADBCBXAXECXX" "XXBXDEDXEEDCAEZA  
DBCXAXECXX" "XXBXDEDXEEDCAEZADBCBXAXECXX" ...  
## $ DS_VT_ESC_OFG : chr "CACDDDD" "CAADEDDE" "CAAACBBE" "BAADBCC" ...  
## $ DS_VT_ACE_OFG : int 11100010 11001011 11000101 1000100 11010101 10001100 11000101 11100011 100011  
1 1111101 ...  
## $ DS_VT_ESC_OCE : chr "BADBDADAEEACAABAADBEBCBDACCCA" "BBDBEEDAEBCEAEABADBCBBEAECCA" "CDDBCDAEEDCACDA  
DDCBBDAAECCA" "BACAAADBEBEACDDBCACBAABACBC" ...  
## $ DS_VT_ACE_OCE : num 9.91e+26 9.91e+26 9.91e+26 9.91e+26 9.91e+26 ...  
## $ TP_PRES : int 555 555 555 555 555 555 555 555 555 555 ...  
## $ TP_PR_GER : int 555 555 555 555 555 555 555 555 555 555 ...  
## $ TP_PR_OB_FG : int 555 555 555 555 555 555 555 555 555 555 ...  
## $ TP_PR_DI_FG : int 555 555 555 555 555 555 555 555 555 333 555 ...  
## $ TP_PR_OB_CE : int 555 555 555 555 555 555 555 555 555 555 555 ...  
## $ TP_PR_DI_CE : int 555 555 555 555 555 555 555 555 333 555 555 ...  
## $ TP_SFG_D1 : int 555 555 555 336 555 555 555 555 333 555 ...  
## $ TP_SFG_D2 : int 555 555 555 333 555 333 555 555 333 555 ...  
## $ TP_SCE_D1 : int 333 555 555 333 555 555 555 333 336 555 ...  
## $ TP_SCE_D2 : int 333 555 555 555 555 555 555 333 336 555 ...  
## $ TP_SCE_D3 : int 555 555 555 333 555 333 555 333 336 555 ...  
## $ NT_GER : num 42.8 59.3 60.9 25 42.4 40 55.1 38.7 50 50.6 ...  
## $ NT_FG : num 50.6 57.9 47.2 15 58.1 26.3 39.2 55.5 30 59.8 ...  
## $ NT_OBJ_FG : num 50 62.5 50 25 62.5 37.5 50 62.5 50 75 ...  
## $ NT_DIS_FG : num 51.5 51 43 0 51.5 9.5 23 45 0 37 ...  
## $ NT_FG_D1 : int 63 55 33 0 42 19 36 47 0 48 ...  
## $ NT_FG_D1_PT : int 75 75 65 0 50 55 60 55 0 60 ...  
## $ NT_FG_D1_CT : int 60 50 25 0 40 10 30 45 0 45 ...  
## $ NT_FG_D2 : int 40 47 53 0 61 0 10 43 0 26 ...  
## $ NT_FG_D2_PT : int 60 55 65 0 65 0 50 75 0 50 ...  
## $ NT_FG_D2_CT : int 35 45 50 0 60 0 0 35 0 20 ...  
## $ NT_CE : num 40.2 59.7 65.4 28.3 37.1 44.5 60.4 33.1 56.7 47.5 ...  
## $ NT_OBJ_CE : num 44.4 66.7 72.2 33.3 38.9 50 61.1 38.9 66.7 50 ...  
## $ NT_DIS_CE : num 16.7 20 26.7 0 26.7 13.3 56.7 0 0 33.3 ...  
## $ NT_CE_D1 : int 0 40 40 0 40 40 40 0 0 30 ...  
## $ NT_CE_D2 : int 0 20 40 0 40 0 80 0 0 20 ...  
## $ NT_CE_D3 : int 50 0 0 0 0 0 50 0 0 50 ...  
## $ CO_RS_I1 : chr "D" "B" "C" "D" ...  
## $ CO_RS_I2 : chr "D" "C" "C" "C" ...  
## $ CO_RS_I3 : chr "C" "A" "C" "E" ...  
## $ CO_RS_I4 : chr "A" "A" "B" "B" ...  
## $ CO_RS_I5 : chr "A" "C" "B" "B" ...  
## $ CO_RS_I6 : chr "C" "B" "C" "B" ...  
## $ CO_RS_I7 : chr "A" "B" "B" "B" ...  
## $ CO_RS_I8 : chr "D" "D" "D" "D" ...  
## $ CO_RS_I9 : chr "D" "D" "D" "E" ...  
## $ QE_I01 : chr "A" "B" "A" "A" ...  
## $ QE_I02 : chr "D" "A" "A" "A" ...  
## $ QE_I03 : chr "A" "A" "A" "A" ...  
## $ QE_I04 : chr "D" "B" "D" "D" ...  
## $ QE_I05 : chr "D" "B" "D" "D" ...  
## $ QE_I06 : chr "B" "C" "B" "B" ...  
## $ QE_I07 : chr "E" "D" "E" "D" ...  
## $ QE_I08 : chr "C" "D" "E" "D" ...  
## $ QE_I09 : chr "B" "E" "D" "D" ...  
## $ QE_I10 : chr "A" "E" "E" "A" ...  
## $ QE_I11 : chr "H" "H" "H" "H" ...  
## $ QE_I12 : chr "A" "A" "A" "A" ...  
## $ QE_I13 : chr "D" "D" "B" "B" ...  
## $ QE_I14 : chr "A" "F" "A" "A" ...  
## $ QE_I15 : chr "A" "A" "A" "A" ...  
## $ QE_I16 : int 50 35 35 35 35 35 35 35 35 35 ...  
## $ QE_I17 : chr "A" "A" "A" "A" ...  
## $ QE_I18 : chr "A" "C" "A" "A" ...  
## $ QE_I19 : chr "B" "C" "B" "B" ...  
## $ QE_I20 : chr "I" "H" "A" "A" ...  
## $ QE_I21 : chr "A" "A" "B" "A" ...  
## $ QE_I22 : chr "B" "C" "D" "E" ...  
## $ QE_I23 : chr "B" "C" "D" "E" ...  
## $ QE_I24 : chr "A" "A" "B" "A" ...  
## $ QE_I25 : chr "C" "A" "C" "A" ...  
## $ QE_I26 : chr "" "" "" "C" ...  
## $ QE_I27 : int 6 6 6 6 6 6 6 6 6 6 ...  
## $ QE_I28 : int 6 6 6 6 6 6 6 6 6 6 ...  
## $ OF_T29 : int 6 6 6 6 6 6 6 6 6 6
```

```
##  ~ QE_I20          . int  0 0 0 0 0 0 0 0 0 ...
##  $ QE_I30          : int  6 6 6 6 6 6 6 6 6 ...
##  [list output truncated]
```

```

#Transformação , manipulação e limpeza dos dados

# aqui vamos selecionar as variáveis de interesse no nosso conjunto de dados
dados=base_enade%>%dplyr::select(NT_OBJ_FG,
                                CO_GRUPO,
                                CO_REGIAO_CURSO,
                                QE_I02,
                                CO_TURNO_GRADUACAO,CO_UF_CURSO)

View(dados)
# aqui filtrei um curso para analisar nesse caso, foi escolhido ciências econômicas
bd_enade=dados%>%filter(CO_GRUPO==13)
View(bd_enade)
# transformação do código 13 em ciência econômicas
bd_enade=bd_enade%>% mutate (CURSO = case_when (CO_GRUPO == 13 ~ "CIÊNCIAS ECONÔMICAS"))
View(bd_enade)
#nomeando a variável nota
names(bd_enade)[1] = "NOTAS"
View(bd_enade)
# aqui uso a função novamente para transformar e classificar CO_REGIAO_CURSO
bd_enade=bd_enade%>% mutate (REGIAO = case_when (CO_REGIAO_CURSO == 1 ~ "Norte",
                                                  CO_REGIAO_CURSO == 2 ~ "Nordeste",
                                                  CO_REGIAO_CURSO == 3 ~ "Sudeste",
                                                  CO_REGIAO_CURSO == 4 ~ "Sul",
                                                  CO_REGIAO_CURSO == 5 ~ "Centro-Oeste" ))

View(bd_enade)
#transformação e classificação da QE_I02 em seus respectivos nomes
bd_enade=bd_enade%>% mutate (RACA = case_when (QE_I02 == "A" ~ "Branca",
                                                QE_I02 == "B" ~ "Preta",
                                                QE_I02 == "C" ~ "Amarela",
                                                QE_I02 == "D" ~ "Parda",
                                                QE_I02 == "E" ~ "Indígena",
                                                QE_I02 == "F" ~ "Não quero declarar" ))

View(bd_enade)
#transformação de turno
bd_enade=bd_enade%>% mutate (TURNO = case_when (CO_TURNO_GRADUACAO == 1 ~ "Matutino",
                                                CO_TURNO_GRADUACAO == 2 ~ "Vespertino",
                                                CO_TURNO_GRADUACAO == 3 ~ "Integral",
                                                CO_TURNO_GRADUACAO == 4 ~ "Noturno" ))

#Transformação da variável cod_uf_curso em nome dos estados
bd_enade=bd_enade%>% mutate(ESTADOS = case_when (CO_UF_CURSO == 11 ~ "RO",
                                                  CO_UF_CURSO == 12 ~ "AC",
                                                  CO_UF_CURSO == 13 ~ "AM",
                                                  CO_UF_CURSO == 14 ~ "RR",
                                                  CO_UF_CURSO == 15 ~ "PA",
                                                  CO_UF_CURSO == 16 ~ "AP",
                                                  CO_UF_CURSO == 17 ~ "TO",
                                                  CO_UF_CURSO == 21 ~ "MA",
                                                  CO_UF_CURSO == 22 ~ "PI",
                                                  CO_UF_CURSO == 23 ~ "CE",
                                                  CO_UF_CURSO == 24 ~ "RN",
                                                  CO_UF_CURSO == 25 ~ "PB",
                                                  CO_UF_CURSO == 26 ~ "PE",
                                                  CO_UF_CURSO == 27 ~ "AL",
                                                  CO_UF_CURSO == 28 ~ "SE",
                                                  CO_UF_CURSO == 29 ~ "BA",
                                                  CO_UF_CURSO == 31 ~ "MG",
                                                  CO_UF_CURSO == 32 ~ "ES",
                                                  CO_UF_CURSO == 33 ~ "RJ",
                                                  CO_UF_CURSO == 35 ~ "SP",
                                                  CO_UF_CURSO == 41 ~ "PR",
                                                  CO_UF_CURSO == 42 ~ "SC",
                                                  CO_UF_CURSO == 43 ~ "RS",
                                                  CO_UF_CURSO == 50 ~ "MS",
                                                  CO_UF_CURSO == 51 ~ "MT",
                                                  CO_UF_CURSO == 52 ~ "GO",
                                                  CO_UF_CURSO == 53 ~ "DF"
                                                  ))

#excluindo variáveis antigas da base nesse caso , das colunas
bd_enade=bd_enade[, -c(2,3,4,5)]

```

Análise descritiva dos dados

Após uma limpeza e transformação dos dados já podemos trabalhar na análise descritiva

```
#aqui usamos a função describe para se ter uma compreensão geral dos dados sobre cada variavel
```

```
describe(bd_enade)
```

```

## bd_enade
##
## 7 Variables          9582 Observations
## -----
## NOTAS
##      n missing distinct      Info      Mean      Gmd
##    8073    1509         9    0.975    63.68    26.04
##
## lowest :   0.0  12.5  25.0  37.5  50.0, highest:  50.0  62.5  75.0  87.5 100.0
##
## Value          0.0  12.5  25.0  37.5  50.0  62.5  75.0  87.5 100.0
## Frequency        69   231   469   892  1255  1430  1671  1363   693
## Proportion 0.009 0.029 0.058 0.110 0.155 0.177 0.207 0.169 0.086
## -----
## CO_UF_CURSO
##      n missing distinct      Info      Mean      Gmd      .05      .10
##    9582         0      25    0.973    33.59    8.66      21      23
##      .25      .50      .75      .90      .95
##      31      35      35      43      50
##
## lowest : 12 13 14 15 17, highest: 43 50 51 52 53
## -----
## CURSO
##              n              missing              distinct              value
##            9582                  0                  1 CIÊNCIAS ECONÔMICAS
##
## Value      CIÊNCIAS ECONÔMICAS
## Frequency                9582
## Proportion                1
## -----
## REGIAO
##      n missing distinct
##    9582         0         5
##
## lowest : Centro-Oeste Nordeste      Norte      Sudeste      Sul
## highest: Centro-Oeste Nordeste      Norte      Sudeste      Sul
##
## Value      Centro-Oeste      Nordeste      Norte      Sudeste      Sul
## Frequency        512        1979        404        4916        1771
## Proportion      0.053      0.207      0.042      0.513      0.185
## -----
## RACA
##      n missing distinct
##    8495    1087         6
##
## lowest : Amarela      Branca      Indígena      Não quero declarar Parda
## highest: Branca      Indígena      Não quero declarar Parda      Preta
##
## Value          Amarela      Branca      Indígena
## Frequency        199      5187      19
## Proportion      0.023      0.611      0.002
##
## Value      Não quero declarar      Parda      Preta
## Frequency        190      2245      655
## Proportion      0.022      0.264      0.077
## -----
## TURNO
##      n missing distinct
##    9582         0         4
##
## Value      Integral      Matutino      Noturno Vespertino
## Frequency        2219      1369      5844      150
## Proportion      0.232      0.143      0.610      0.016
## -----
## ESTADOS
##      n missing distinct
##    9582         0      25
##
## lowest : AC AL AM BA CE, highest: RS sc SE SP TO
## -----

```



```
#resumo básico do nosso conjunto de dados
summary(bd_enade)
```

```
##      NOTAS      CO_UF_CURSO      CURSO      REGIAO
## Min.   : 0.00   Min.   :12.00   Length:9582   Length:9582
## 1st Qu.: 50.00   1st Qu.:31.00   Class :character   Class :character
## Median : 62.50   Median :35.00   Mode  :character   Mode  :character
## Mean   : 63.68   Mean   :33.59
## 3rd Qu.: 87.50   3rd Qu.:35.00
## Max.   :100.00   Max.   :53.00
## NA's    :1509
##      RACA      TURNO      ESTADOS
## Length:9582   Length:9582   Length:9582
## Class :character   Class :character   Class :character
## Mode  :character   Mode  :character   Mode  :character
##
##
##
##
```

```
#se tiver valores faltantes vamos exclui-los com os seguinte comando
bd_enade=bd_enade[>%na.omit()]
summary(bd_enade) #foi removido todos valores faltantes
```

```
##      NOTAS      CO_UF_CURSO      CURSO      REGIAO
## Min.   : 0.00   Min.   :12.00   Length:8024   Length:8024
## 1st Qu.: 50.00   1st Qu.:29.00   Class :character   Class :character
## Median : 62.50   Median :35.00   Mode  :character   Mode  :character
## Mean   : 63.73   Mean   :33.49
## 3rd Qu.: 87.50   3rd Qu.:35.00
## Max.   :100.00   Max.   :53.00
##      RACA      TURNO      ESTADOS
## Length:8024   Length:8024   Length:8024
## Class :character   Class :character   Class :character
## Mode  :character   Mode  :character   Mode  :character
##
##
##
```

```
#mediana das notas do enade
mediana_enade=median(bd_enade$NOTAS)
mediana_enade
```

```
## [1] 62.5
```

```
#media das notas
media_notas=mean(bd_enade$NOTAS)
media_notas
```

```
## [1] 63.72601
```

```
freq=table(bd_enade$NOTAS)
#valor maximo de notas
max(freq)
```

```
## [1] 1663
```

```
#valor minimo
min(freq)
```

```
## [1] 68
```

```
# verificando a moda
valor_maximo=max(freq)
#frequencia das notas
m=names(freq)
#moda das notas
moda=m[freq==valor_maximo]
moda=as.numeric(moda)
moda
```

```
## [1] 75
```

```
#calculando a amplitude das notas
amplitude_notas=max(bd_enade$NOTAS)-min(bd_enade$NOTAS)
amplitude_notas
```

```
## [1] 100
```

```
#apresentando a variância da notas
variancia=var(bd_enade$NOTAS)
variancia
```

```
## [1] 536.7344
```

```
#desvio padrão
desvio=sd(bd_enade$NOTAS)
desvio
```

```
## [1] 23.16753
```

```
#calculando o coeficiente de variação
CV=desvio/media_notas
CV
```

```
## [1] 0.363549
```

```
#curtose dos dados
curtose=kurtosis(bd_enade$NOTAS)
curtose
```

```
## [1] -0.4614936
```

```
#assimetria dos dados
ASM= skewness(bd_enade$NOTAS)
ASM
```

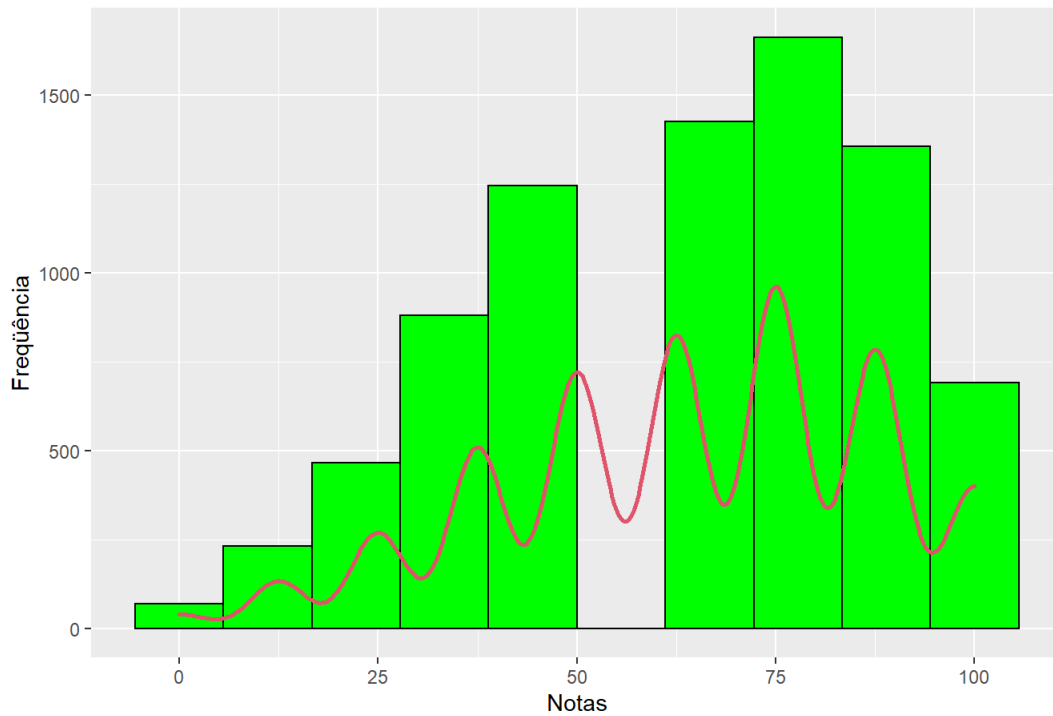
```
## [1] -0.4112657
```

```
resumo_geral=c(media_notas,mediana_enade,amplitude_notas,variancia,desvio,CV,curtose,ASM,moda)
resumo_geral
```

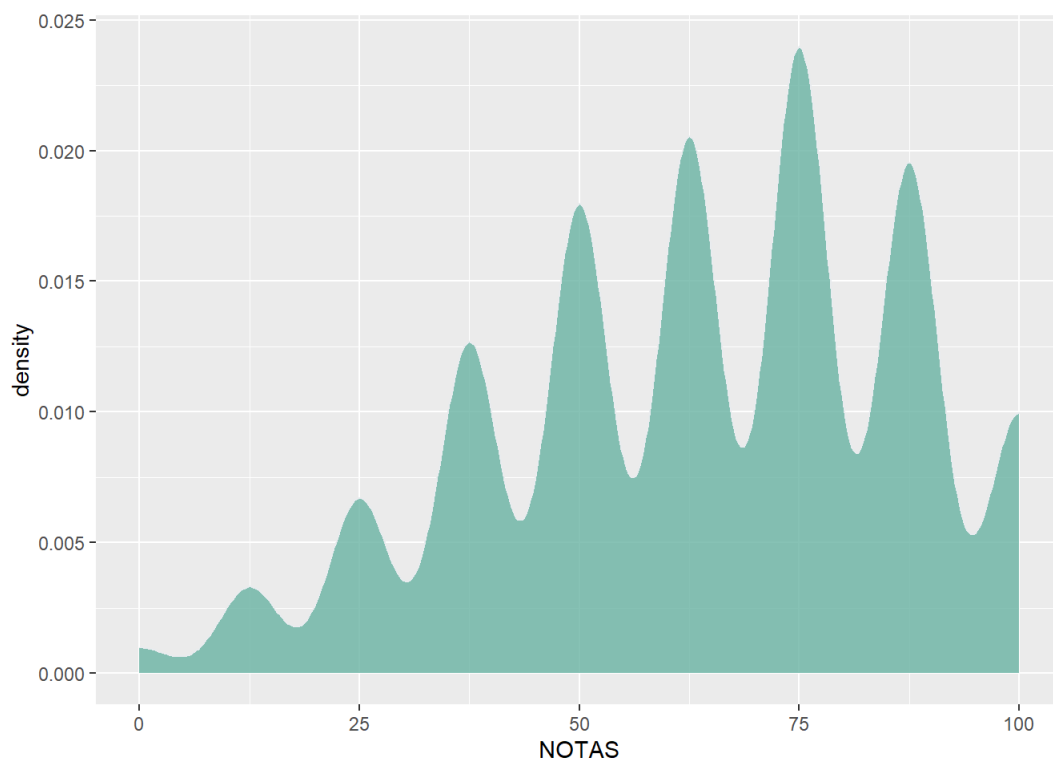
```
## [1] 63.7260095 62.5000000 100.0000000 536.7344362 23.1675298 0.3635490
## [7] -0.4614936 -0.4112657 75.0000000
```

```
hist_notas=ggplot(bd_enade, aes(x = NOTAS)) +
  geom_histogram(color = "black", fill = "green", bins = 10) +
  geom_density(col = 2, size = 1, aes(y = 5 * ..count..)) +
  labs(title="Histograma e Curva de Densidade das Notas dos Alunos de Ciências Econômicas")+
  labs(x = "Notas", y = "Frequência")
hist_notas
```

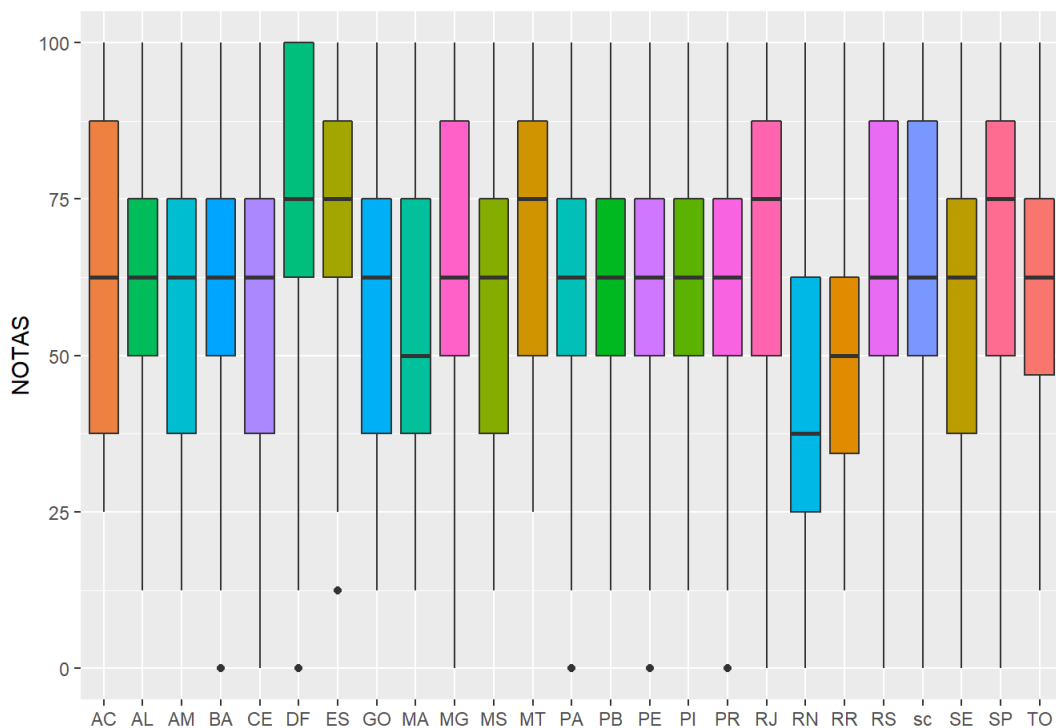
Histograma e Curva de Densidade das Notas dos Alunos de Ciências Econômica



```
#grafico de densidade das notas
bd_enade%>%
  filter( NOTAS<200 )%>%
  ggplot( aes(x=NOTAS,)) +
  geom_density(fill="#69b3a2", color="#e9ecef", alpha=0.8)
```



```
#grafico de boxplot de notas x estados
bd_enade %>%
  mutate(class = fct_reorder(ESTADOS, NOTAS, .fun='length' )) %>%
  ggplot( aes(x=ESTADOS, y=NOTAS, fill=class)) +
  geom_boxplot() +
  xlab("ESTADOS") +
  theme(legend.position="none") +
  xlab("") +
  xlab("")
```



```
#verificação da frequência e proporção TURNO
table(bd_enade$TURNO)
```

```
##
##      Integral      Matutino      Noturno Vespertino
##      1944          1175          4772          133
```

```
prop.table(table(bd_enade$TURNO))
```

```
##
##      Integral      Matutino      Noturno Vespertino
## 0.24227318 0.14643569 0.59471585 0.01657527
```

```
#Verificação da frequência e proporção TURNO X REGIÃO
table(bd_enade$TURNO,bd_enade$REGIAO)
```

```
##
##              Centro-Oeste Nordeste Norte Sudeste  Sul
##      Integral           123      201      11    1402  207
##      Matutino           42       311      89     545  188
##      Noturno           253     1068     231    2161 1059
##      Vespertino          0        82      14       37   0
```

```
prop.table(table(bd_enade$TURNO,bd_enade$REGIAO))
```

```
##
##              Centro-Oeste      Nordeste      Norte      Sudeste      Sul
##      Integral  0.015329013 0.025049850 0.001370887 0.174725823 0.025797607
##      Matutino  0.005234297 0.038758724 0.011091725 0.067921236 0.023429711
##      Noturno   0.031530409 0.133100698 0.028788634 0.269317049 0.131979063
##      Vespertino 0.000000000 0.010219342 0.001744766 0.004611167 0.000000000
```

```
#verificando a freq e prop raça x regioa
table(bd_enade$RACA,bd_enade$REGIAO)
```

```
##
##          Centro-Oeste Nordeste Norte Sudeste Sul
## Amarela          20      43      6      92  25
## Branca           195     635     95    2806 1163
## Indígena          0       5       4       7   3
## Não quero declarar 3      43      3      91  37
## Parda            150     760    204     843 173
## Preta             50     176     33     306  53
```

```
prop.table(table(bd_enade$RACA,bd_enade$REGIAO))
```

```
##
##          Centro-Oeste      Nordeste      Norte      Sudeste
## Amarela      0.0024925224 0.0053589232 0.0007477567 0.0114656032
## Branca       0.0243020937 0.0791375872 0.0118394816 0.3497008973
## Indígena     0.0000000000 0.0006231306 0.0004985045 0.0008723829
## Não quero declarar 0.0003738784 0.0053589232 0.0003738784 0.0113409771
## Parda        0.0186939182 0.0947158524 0.0254237288 0.1050598205
## Preta        0.0062313061 0.0219341974 0.0041126620 0.0381355932
##
##          Sul
## Amarela      0.0031156530
## Branca       0.1449401795
## Indígena     0.0003738784
## Não quero declarar 0.0046111665
## Parda        0.0215603190
## Preta        0.0066051844
```

```
#agregação turno e notas
```

```
NOTAS_TURNO=bd_enade%>%select(TURNO,NOTAS)%>%
  group_by (TURNO) %>%
  summarise (MEDIA=mean(NOTAS))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
# agregando por turno e regioao
NOTA_TUR_REG = bd_enade %>% select (TURNO, REGIAO, NOTAS) %>%
  group_by (TURNO, REGIAO) %>%
  summarise (MEDIA = mean(NOTAS))
```

```
## `summarise()` regrouping output by 'TURNO' (override with `.groups` argument)
```

```
# agregando aqui por turno , regioao e notas
NTR = bd_enade%>%select(TURNO,REGIAO,NOTAS)%>%
  group_by(TURNO,REGIAO)%>%
  summarise (MEDIA=mean(NOTAS))
```

```
## `summarise()` regrouping output by 'TURNO' (override with `.groups` argument)
```

```
NTR
```

```
## # A tibble: 18 x 3
## # Groups:   TURNO [4]
##   TURNO      REGIAO      MEDIA
##   <chr>    <chr>    <dbl>
## 1 Integral Centro-Oeste  74.0
## 2 Integral Nordeste    63.6
## 3 Integral Norte      68.2
## 4 Integral Sudeste     68.8
## 5 Integral Sul        66.4
## 6 Matutino Centro-Oeste  70.2
## 7 Matutino Nordeste    61.5
## 8 Matutino Norte      61.5
## 9 Matutino Sudeste     71.4
## 10 Matutino Sul        65.7
## 11 Noturno  Centro-Oeste  55.5
## 12 Noturno  Nordeste    57.9
## 13 Noturno  Norte      58.5
## 14 Noturno  Sudeste     63.9
## 15 Noturno  Sul        61.0
## 16 Vespertino Nordeste    57.6
## 17 Vespertino Norte      59.8
## 18 Vespertino Sudeste     58.8
```

```
# agregando os estados com as médias das notas
NOTAS_ESTADOS=bd_enade%>%select (ESTADOS,NOTAS)%>%
  group_by (ESTADOS) %>%
  summarise (MEDIA=mean (NOTAS) )
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
NOTAS_ESTADOS
```

```
## # A tibble: 25 x 2
##   ESTADOS MEDIA
##   <chr>    <dbl>
## 1 AC      64.1
## 2 AL      57.3
## 3 AM      58.2
## 4 BA      64.5
## 5 CE      57.9
## 6 DF      74.8
## 7 ES      73.4
## 8 GO      54.9
## 9 MA      56.5
## 10 MG     64.7
## # ... with 15 more rows
```

Comentários finais

Analisando as regiões por turno e media de notas , a região centro-oeste tem uma maior média em comparação com as demais no turno integral , em seguida o sudeste também apresenta melhor desempenho médio de notas , enquanto o nordeste tem uma das menores médias de notas no turno integral. No turno matutino a região sudeste apresentou uma melhor média , as regiões nordeste e norte teve uma média menor em relação as outras , no periodo nortuno mais uma vez a região sudeste apresentou uma média superior as demais regiões , o centro-oeste teve a menor nota do grupo. No periodo vespertino o norte teve uma média superior as demais . Portanto , o turno integral tem a maior nota média (68,3) enquanto o vespertino tem a menor média entre o turno.

Sobre as médias das notas , o estado do Distrito Federal(74.77) e Espirito Santo(73.39) apresentou as notas médias maiores.

Gostaria de deixar como fonte referência e inspiração dessa análise alguns canais no youtube que me ajudaram nessa trilha de aprendizado.

1.[Análise descritiva na Base de dados do ENADE Usando R - Camila Koehler](#)

2.[R Markdown](#)

Estou aberto a sugestões de melhorias e criticas construtivas, esse é um dos primeiros trabalhos que pretendo iniciar no trilheiro de dados , uma verdadeira trilha de experiências e aprendizados.