

# Análise Exploratória de Dados

Luiz Carlos Castro Cunha Junior

29/05/2021



## Estudo de Caso: Limite de Crédito

A base de dados traz informações dos clientes de um banco que solicitaram limite de crédito. Durante a análise deverá ser respondida às seguintes questões de negócio:

- Qual a idade média dos clientes presentes no banco de dados?
- Qual o valor do mínimo, máximo, mediana, Q1 e Q3 da variável idade?
- Existem clientes com idades discrepantes? Analise o boxplot.
- Existem clientes que possuem rendimento total discrepante em relação aos demais clientes? Analise o boxplot.
- A partir de qual valor o rendimento é considerado discrepante?
- A variável rendimento total pode ser considerada simétrica?
- Existem clientes que possuem salário discrepante em relação aos demais clientes?
- A partir de qual valor o salário é considerado discrepante?
- A variável salário pode ser considerada simétrica?
- Existem clientes que possuem limite de cheque especial discrepante em relação aos demais clientes? Analise o boxplot.
- A partir de qual valor o limite de cheque especial é considerado discrepante?
- A variável limite de cheque especial pode ser considerada simétrica?

## Carregando os pacotes a serem utilizados

```
library(tidyverse)
library(readxl)
library(rstatix)
library(moments)
library(knitr)
```

```
library(rmarkdown)
library(kableExtra)
library(emmeans)
library(htmltools)
```

## Importando o Conjunto de Dados

```
dataset<-readxl::read_xlsx("Exercícios.xlsx",
                           sheet = "Base de Dados 1")
colnames(dataset)[3]<-'RendimentoTotal'

knitr::kable(head(dataset,5))
```

Código do Cliente	Idade	RendimentoTotal	Salário	Limite de Crédito Imediato	Limite do Cheque Especial
12	72	4300	4300	2000	1000
44	75	4400	4400	3000	1000
30	66	4800	4800	440	1500
2	35	5000	5000	1000	1000
35	69	5000	5000	2000	2500

## Visualizando o tipo de dados das variáveis

```
str(dataset)
```

```
## tibble [48 x 6] (S3: tbl_df/tbl/data.frame)
##  $ Código do Cliente      : num [1:48] 12 44 30 2 35 39 42 47 5 19 ...
##  $ Idade                  : num [1:48] 72 75 66 35 69 47 68 44 54 56 ...
##  $ RendimentoTotal        : num [1:48] 4300 4400 4800 5000 5000 5000 5000 5800 6000 6000 ...
##  $ Salário                : num [1:48] 4300 4400 4800 5000 5000 5000 5000 5800 6000 6000 ...
##  $ Limite de Crédito Imediato: num [1:48] 2000 3000 440 1000 2000 2000 380 500 1790 2000 ...
##  $ Limite do Cheque Especial : num [1:48] 1000 1000 1500 1000 2500 1700 600 800 3600 2500 ...
```

## Perguntas de Negócio

Qual a idade média dos clientes presentes no banco de dados?

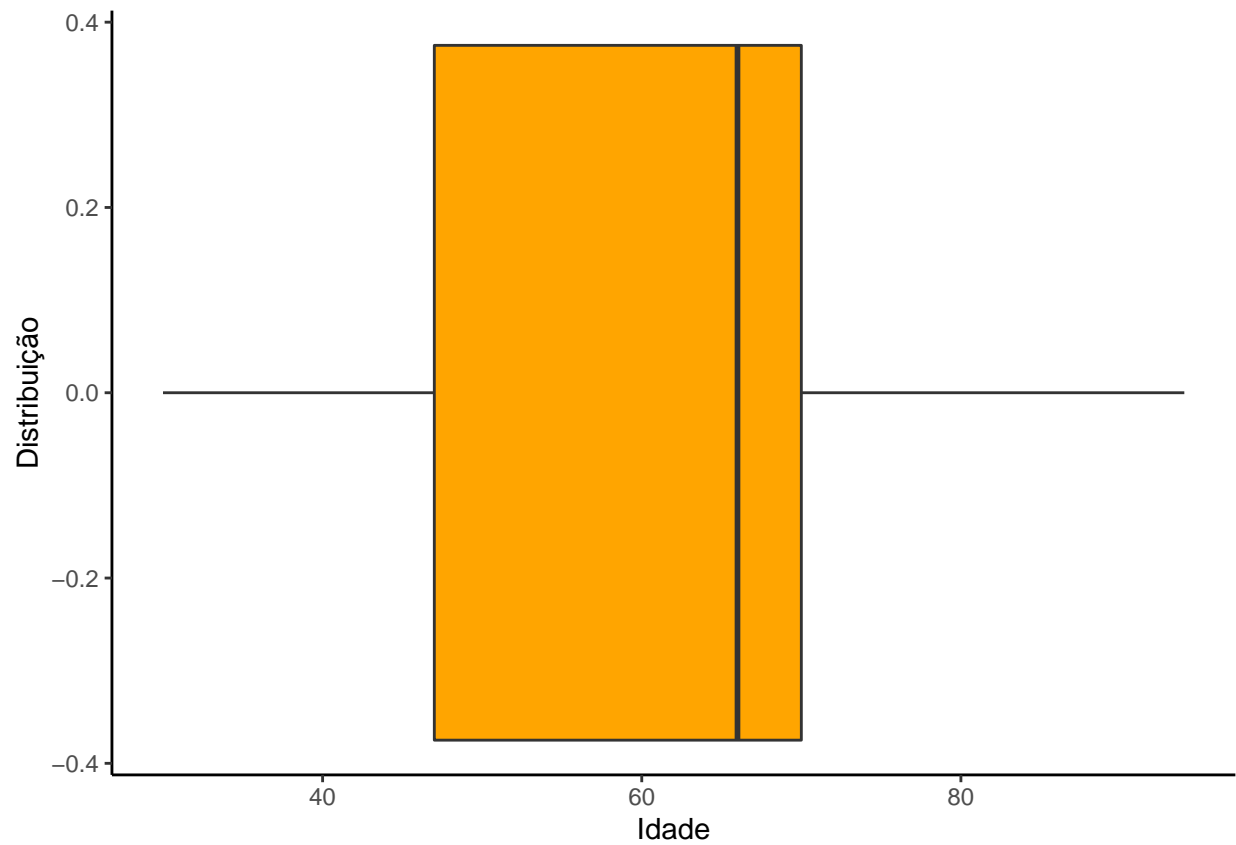
A idade média dos clientes é 60.98

Qual o valor do mínimo, máximo, mediana, Q1 e Q3 da variável idade?

- Idade Mínima:30
- Idade Máxima:47
- Mediana de Idade:30
- Primeiro Quartil:47
- Terceiro Quartil:70

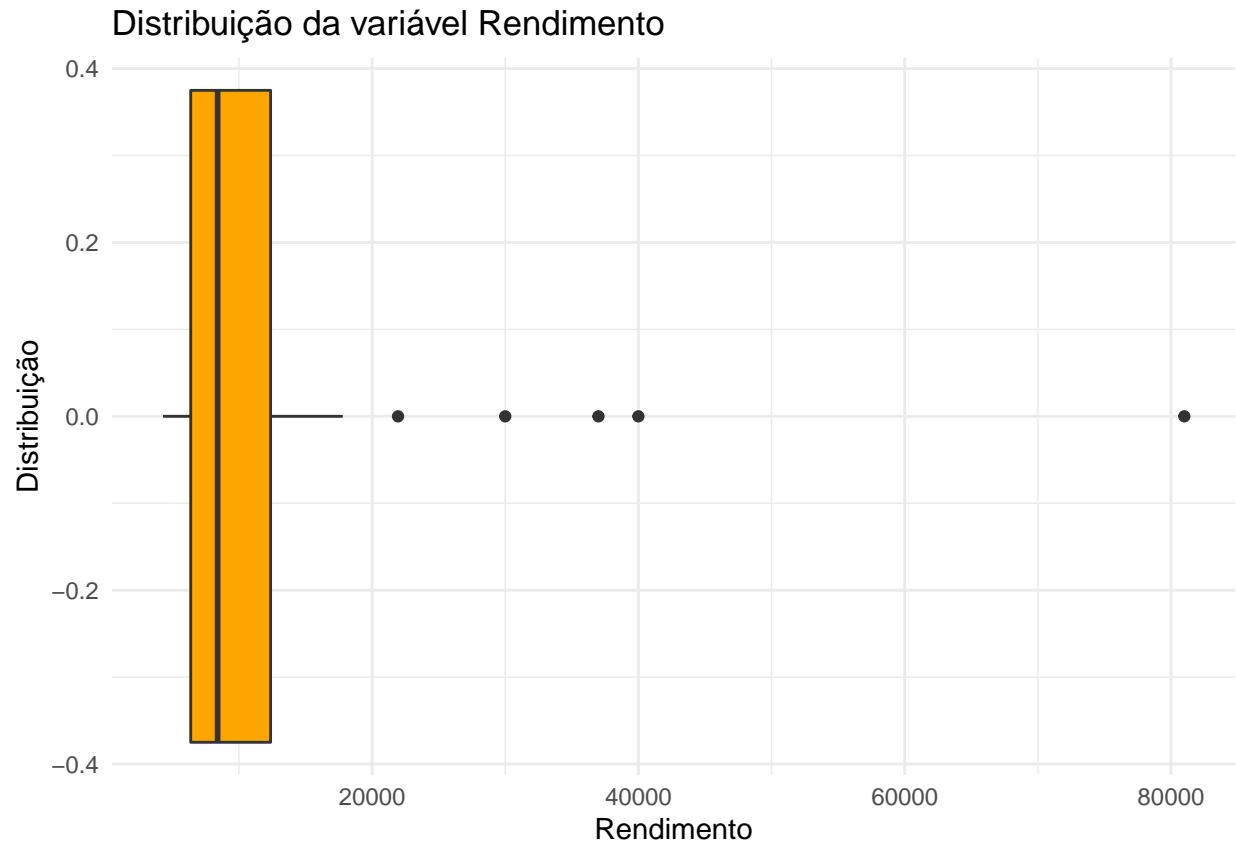
Existem clientes com idades discrepantes?

Conforme podemos analisar no gráfico abaixo *não temos clientes com idades discrepantes* em relação aos outros clientes.



Existem clientes que possuem rendimento total discrepante em relação aos demais clientes?

Podemos observar no gráfico abaixo que possuímos clientes cujo rendimento total é considerado discrepante em relação aos outros clientes.



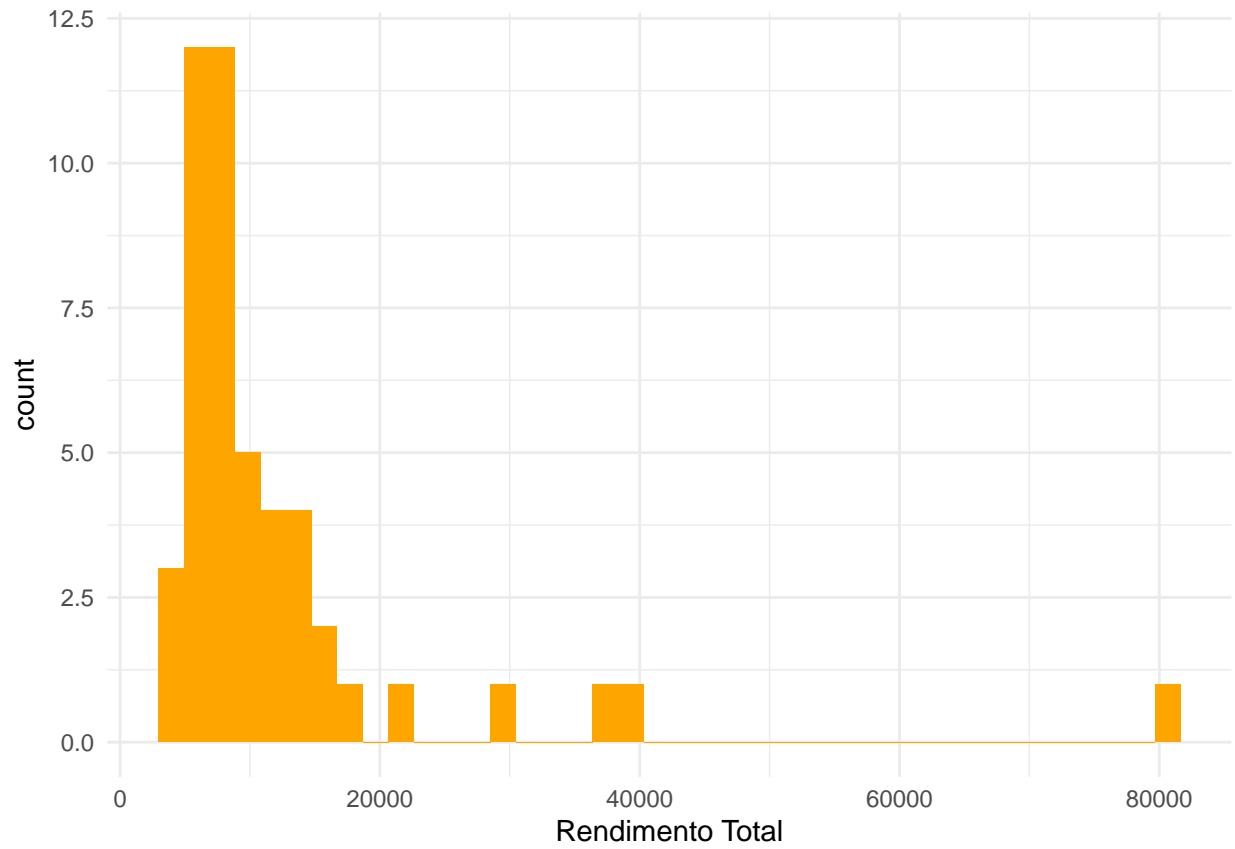
A partir de qual valor o rendimento é considerado discrepante?

A partir do rendimento total de 21375

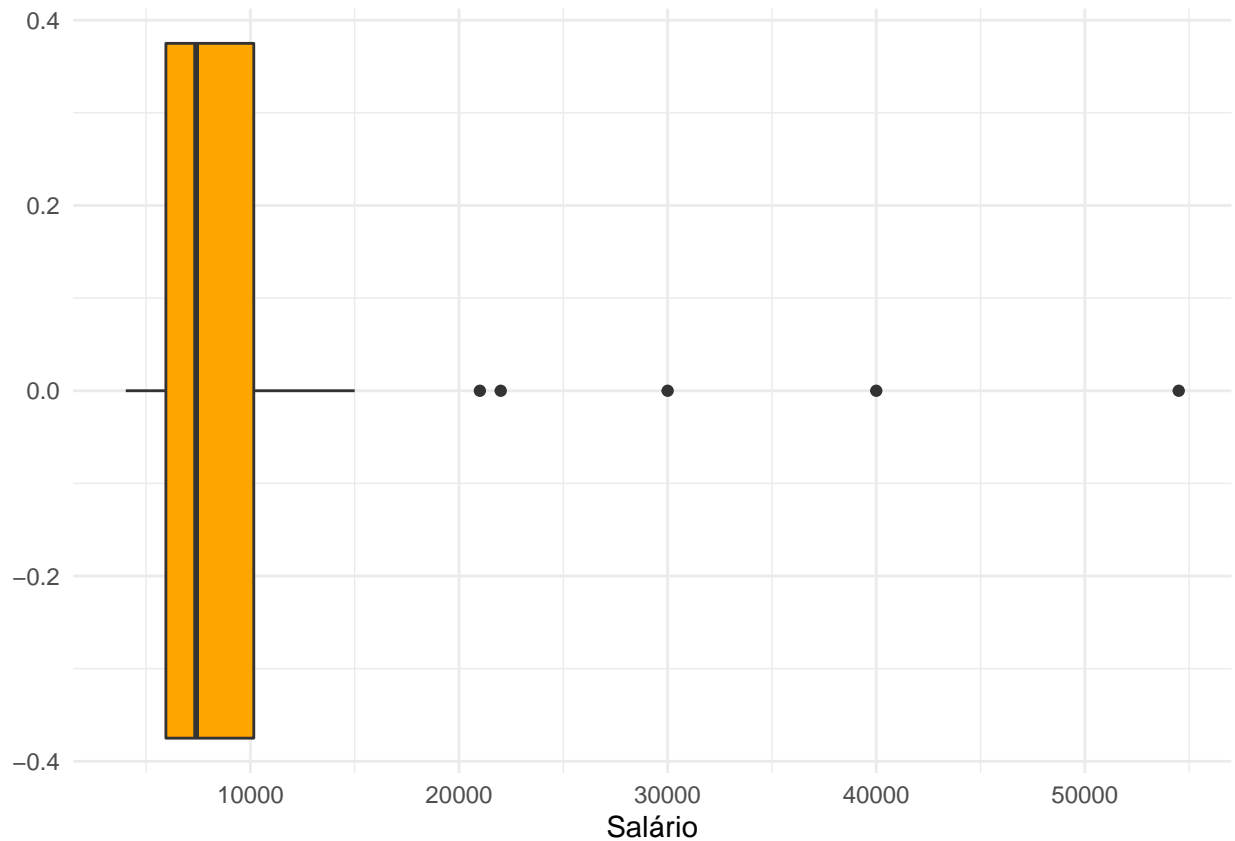
A variável rendimento total pode ser considerada simétrica?

Não, ela é assimétrica para a direita em decorrência de termos a média superior a mediana, conforme podemos observar nos valores abaixo e no histograma.

- Média: 4300
- Mediana: 4800, 4800, 4800, 4800, 4800, 4800
- Coeficiente de assimetria: 3.8524194



Existem clientes que possuem salário discrepante em relação aos demais clientes?



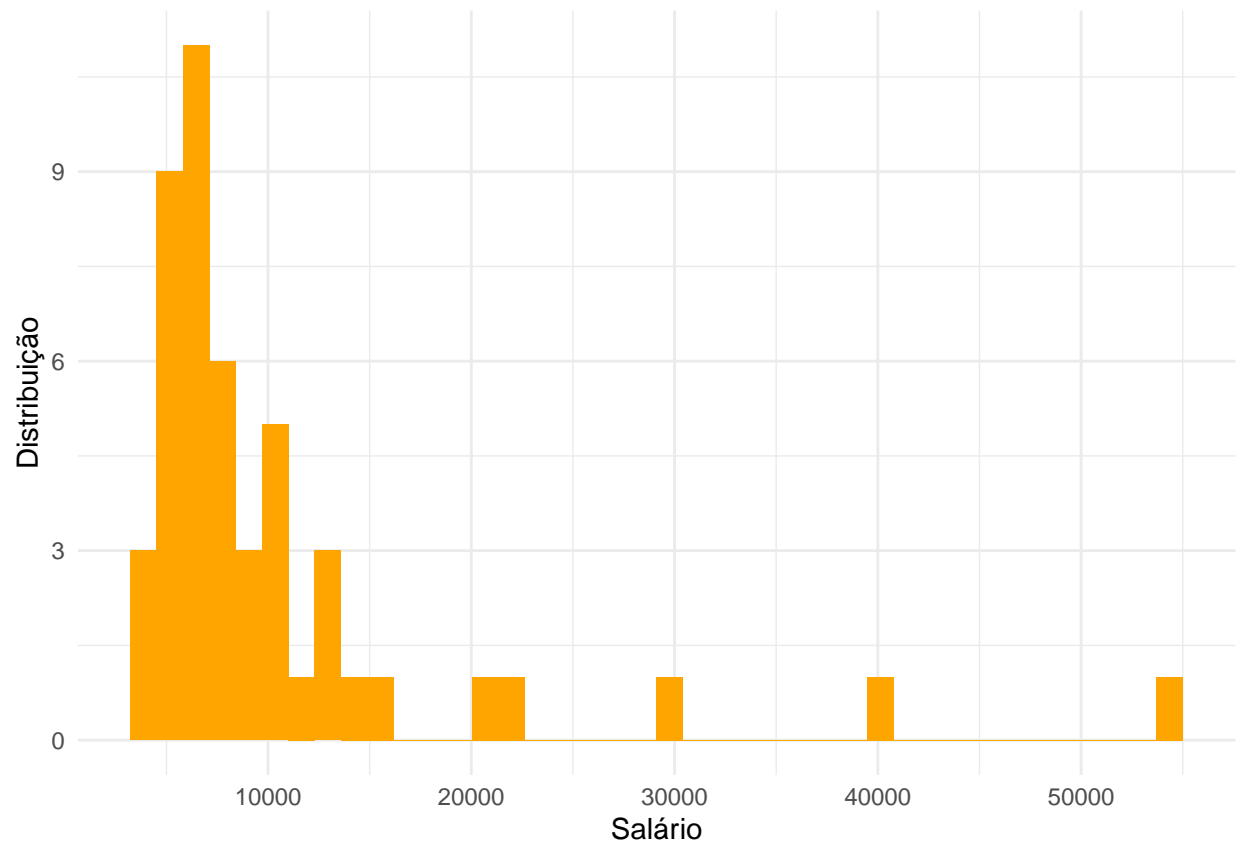
A partir de qual valor o salário é considerado discrepante?

O salário pode ser considerado discrepante a partir de R\$16482.5

A variável salário pode ser considerada simétrica?

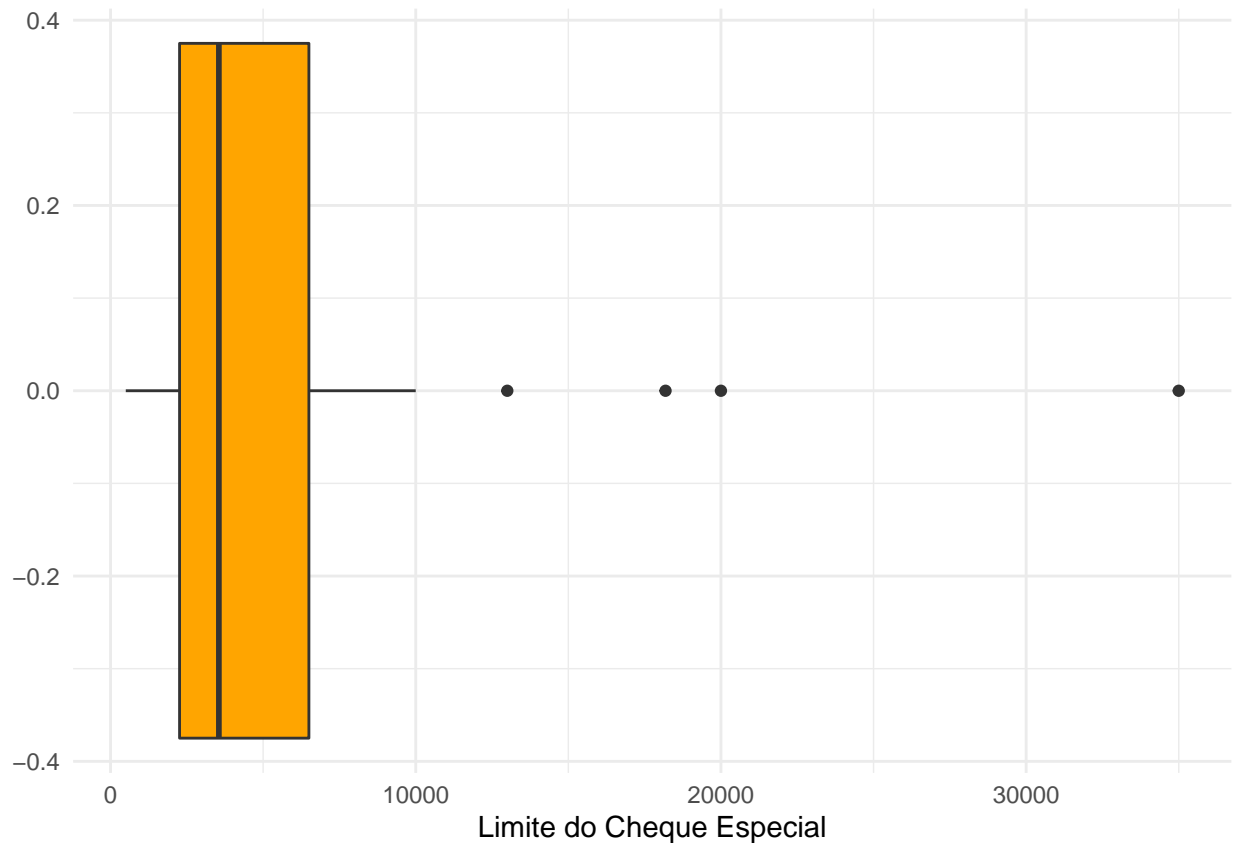
Não. A variável é assimétrica para a direita em decorrência de termos a sua média superior a mediana, conforme podemos observar nos valores abaixo e no histograma.

- Média: 4027
- Mediana: 7400
- Coeficiente de assimetria: 3.1261053



Existem clientes que possuem limite de cheque especial discrepante em relação aos demais clientes?

Sim.



A partir de qual valor o limite de cheque especial é considerado discrepante?

O valor do limite do cheque especial pode ser considerado discrepante a partir de R\$ 12856.25

A variável limite de cheque especial pode ser considerada simétrica?

Não, ela é assimétrica para a direita em decorrência de termos a média superior a mediana, conforme podemos observar nos valores abaixo e no histograma.

- Média: 500
- Mediana: 1500, 1500, 1500, 1500, 1500, 1500
- Coeficiente de assimetria: 2.7935198



