

# ZABBIX 2023 Conference

LATIN AMERICA





**PAULO R. DEOLINDO JUNIOR**  
Zabbix Trainer & DPO na Unirede  
Inteligência em TI

**Conecte-se!**



# Zabbix 6.+ Introdução à Ciência de Dados

# Processo e Objetivos

# Objetivos em Ciência de Dados

- **Análise descritiva**

- O que aconteceu ou está acontecendo no ambiente de dados?

Graphs

Tables

Storytelling

- **Análise diagnóstica**

- Por que algo aconteceu?

Correlation

Discovery

Anomaly

- **Análise preditiva**

- Uso de dados históricos para previsões de padrões e ocorrências futuras

Forecast

Timeleft

Baseline

- **Análise prescritiva**

- Evolução da análise preditiva

Suggestions

# O processo da Ciência de Dados

- **Obter dados**

Coleta

Importação

Geração

- **Suprimir dados**

Pre-processamento

- **Explorar dados**

Funções estatísticas

Funções matemáticas

Funções agregadas

- **Modelar dados**

Classificação

Descoberta de padrões

Aprendizado de máquina

- **Interpretar resultados**

Dashboards

Gráficos

Insights



# Os 4 momentos da distribuição de dados

# Os primeiros 4 momentos de uma distribuição

## 1- Estimativa

# média, mediana, média ponderada, média aparada, 50% percentil, etc.

## 2- Variabilidade

# variância, desvio, desvio padrão, etc.

## 3- Assimetria

# Moderada ou forte, à direita ou à esquerda, Coeficiente de Pearson, etc.

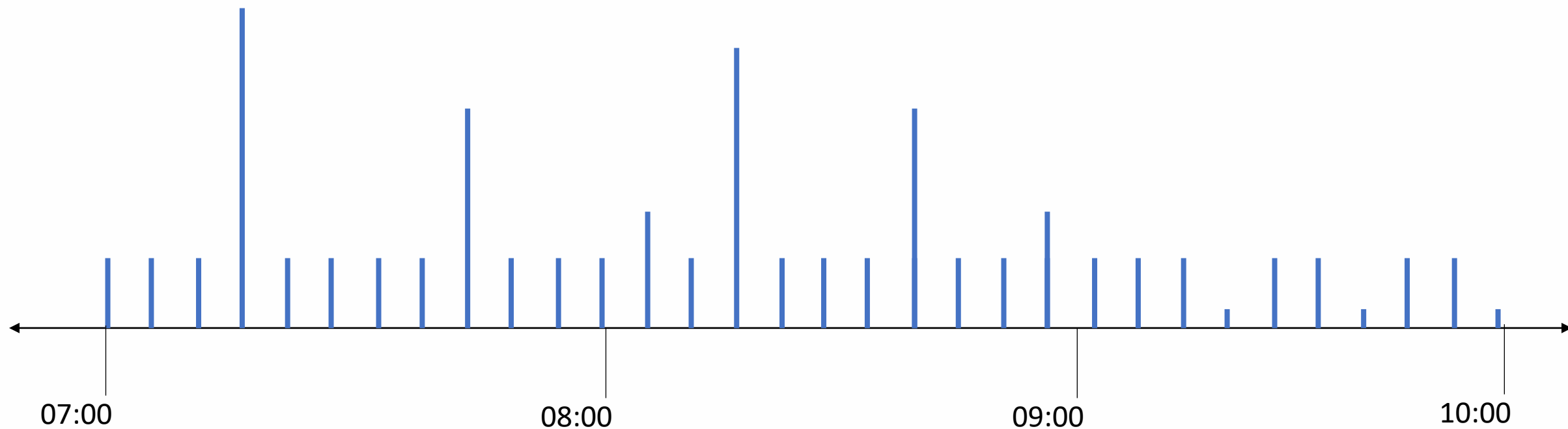
## 4- Curtose

# Tendência a picos o não

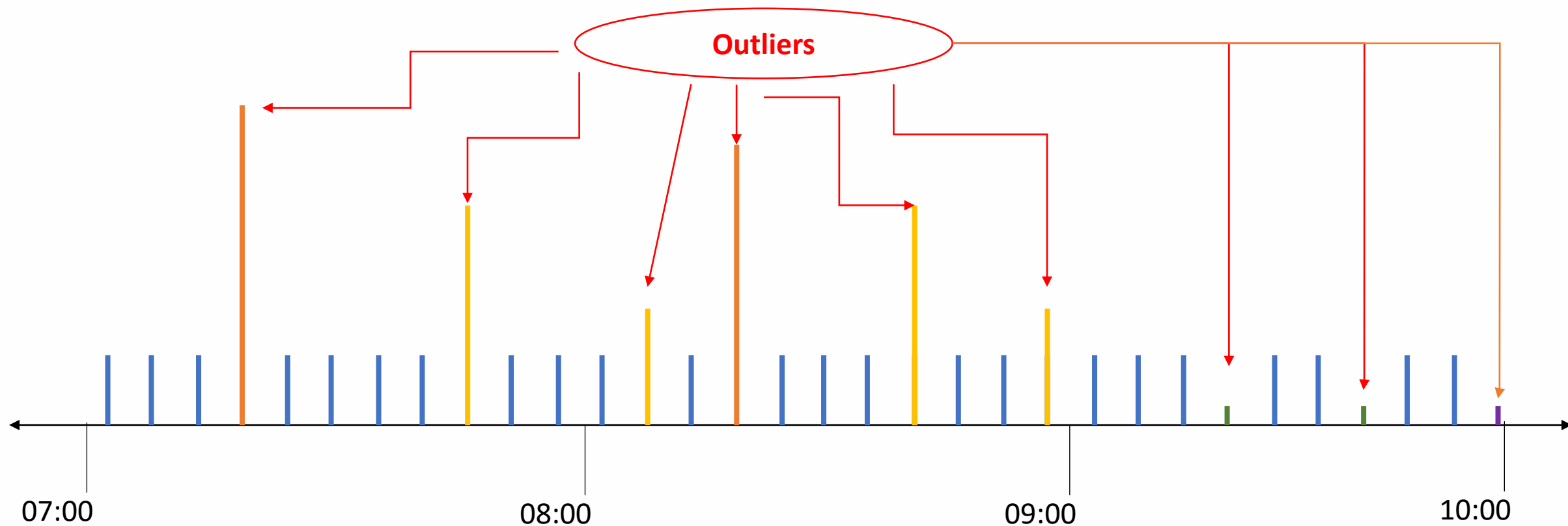


**1º Momento:**  
**Estimativas de localização**

# Distribuição de dados



# Erros | Desvios



# Estimativas de localização

1º

## Média

Média ponderada

## Média móvel ponderada

## Mediana

Mediana ponderada

Média aparada

Robusto

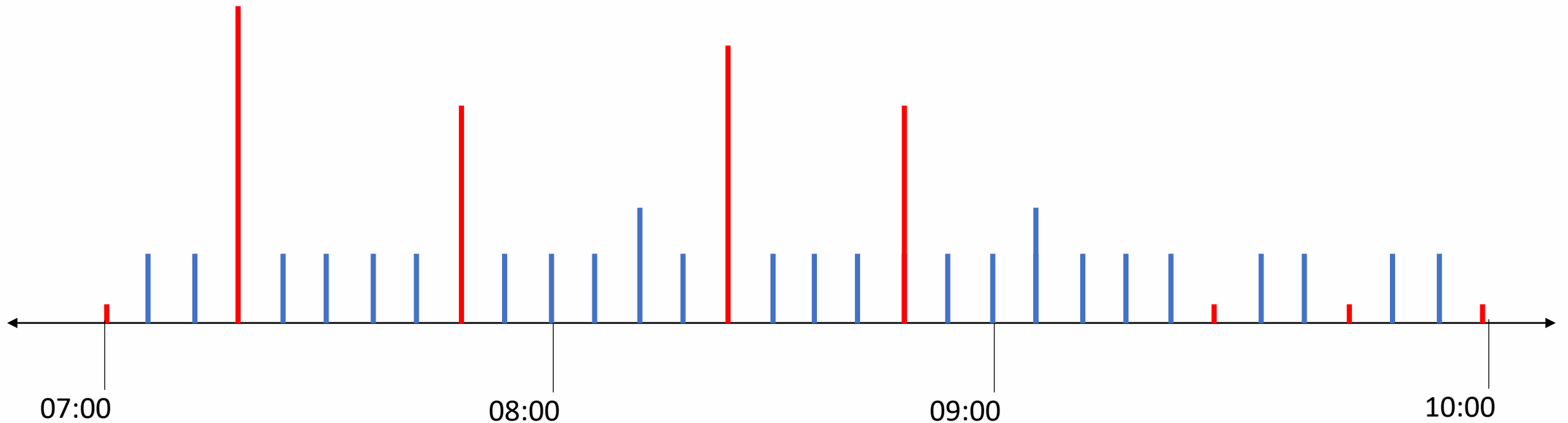
## Outlier

Moda

Percentil

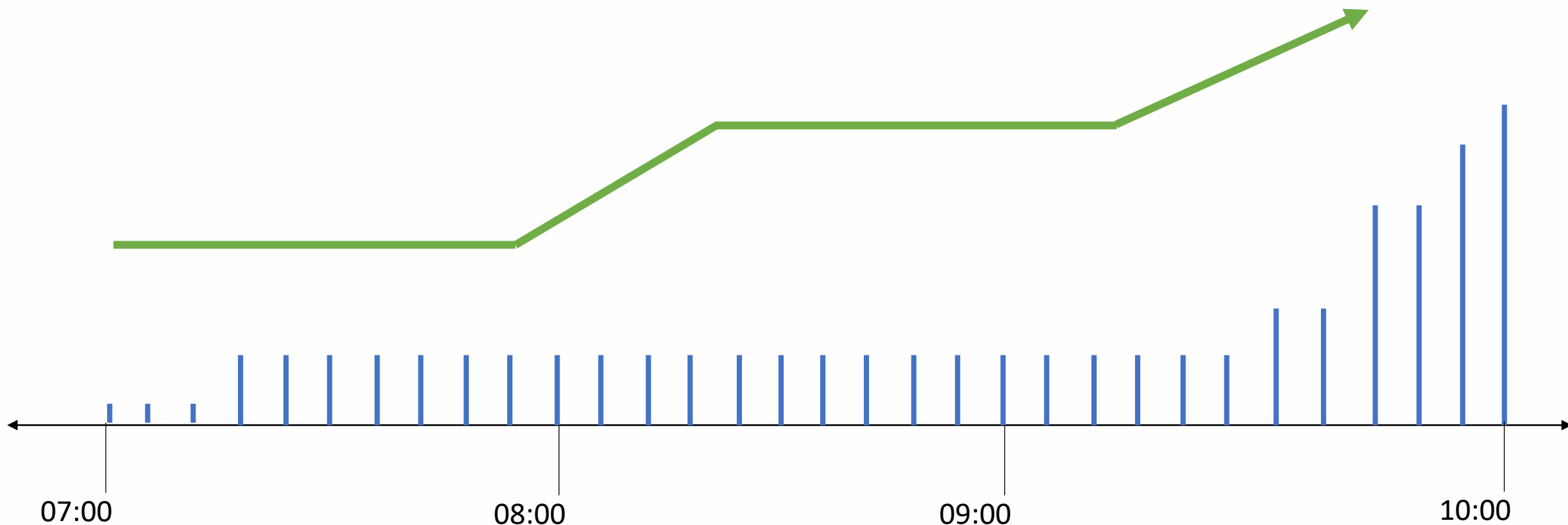
# Estimativas **não** robustas: Média

avg(/host/key,4h)



Ordered by collected time

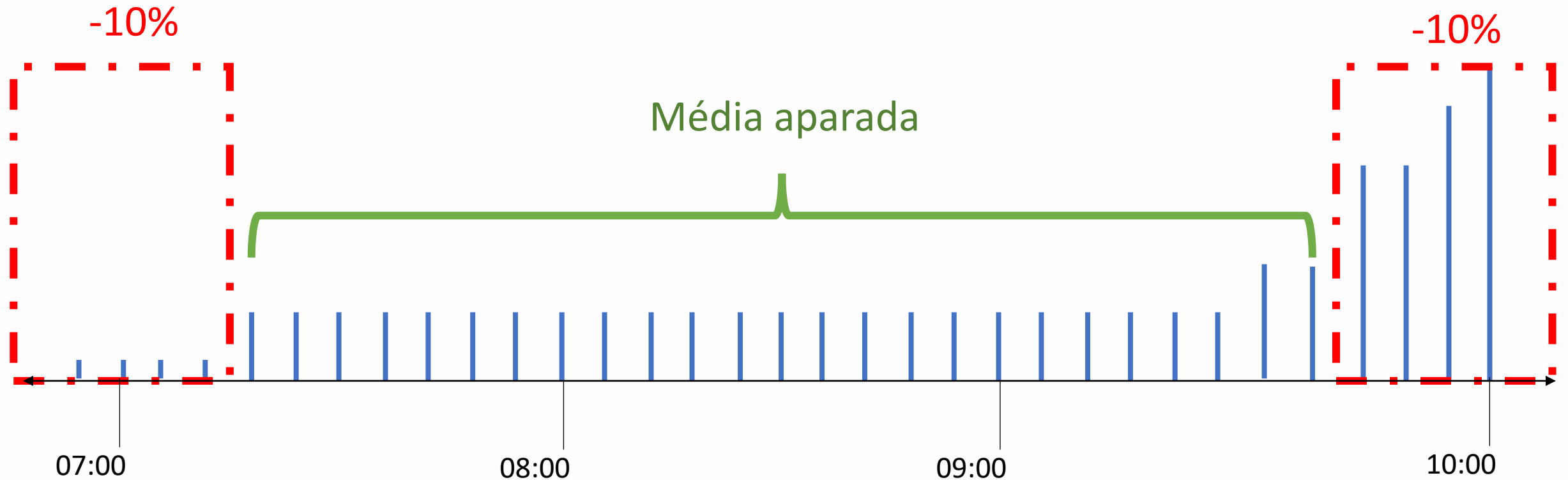
# Classificação: crescente



Ordered by size; not by collected time



# Estimativas robustas: Média aparada

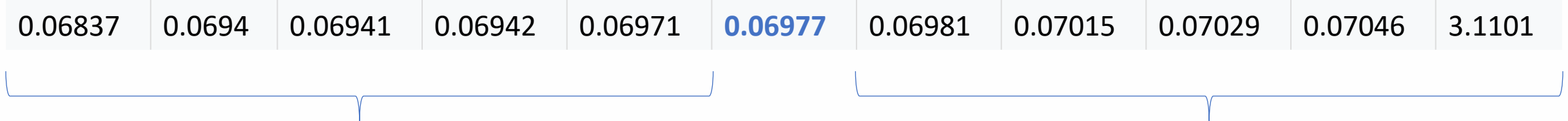




# Estimativas robustas: 50% Percentil x Mediana

Um número **ímpar** de dados

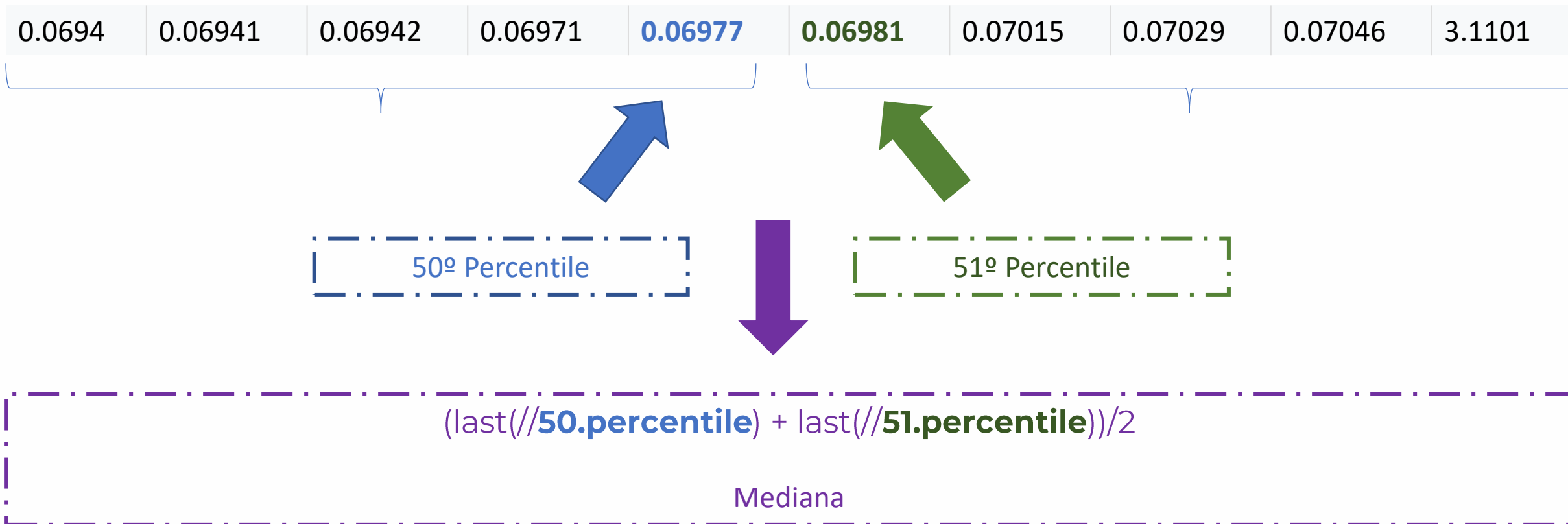
Valor central = Mediana  
50º Percentile



# Estimativas robustas: 50% Percentil x

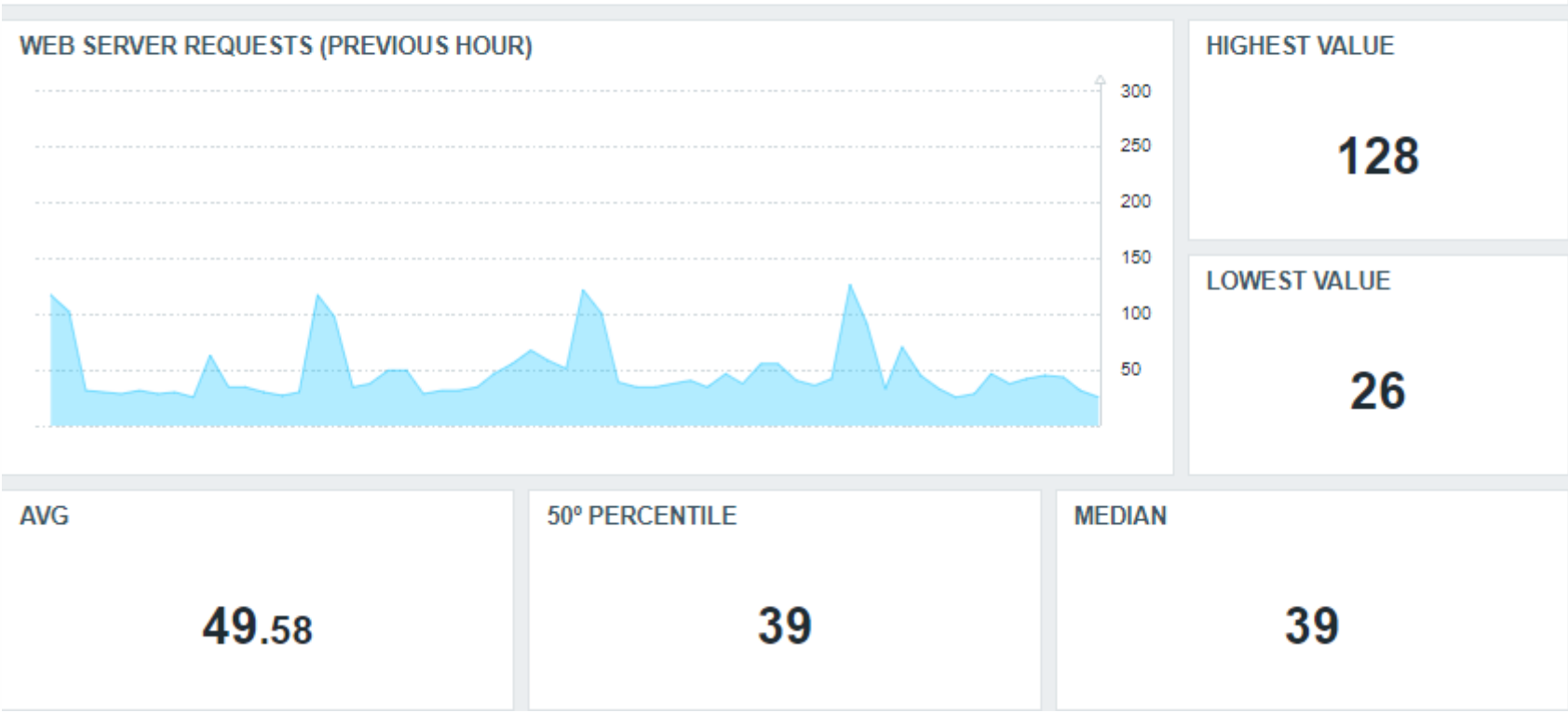
## Mediana

Um número **par** de dados



## LOCATION ESTIMATE

### ABOUT WEB SERVER REQUESTS (P/MINUTE)



# 2º Momento: Variabilidade

# Variabilidade

2º

**Desvio**

**Variância**

**Desvio-padrão**

Desvio absoluto médio

**Desvio absoluto médio da mediana**

**Amplitude**

Estatísticas ordinais

**Percentil**

**Amplitude interquartilica**

# Variabilidade: Amplitude

A **diferença** entre o **maior** e o **menor** valor

<b>0.0694</b>	0.06941	0.06942	0.06971	0.06977	0.06981	0.07015	0.07029	0.07046	<b>3.1101</b>
---------------	---------	---------	---------	---------	---------	---------	---------	---------	---------------

**max**(//key,#10) - **min**(// key,#10)

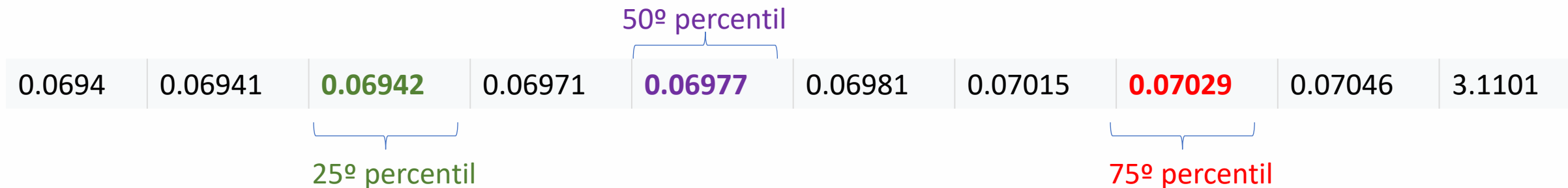
Amplitude



3.0407

# Variabilidade: **IQR** (Amplitude interquartil)

A **diferença** entre o **75º percentil** e **25º percentil** valor



**`last(//percentile.75) - last(//percentile.25)`**

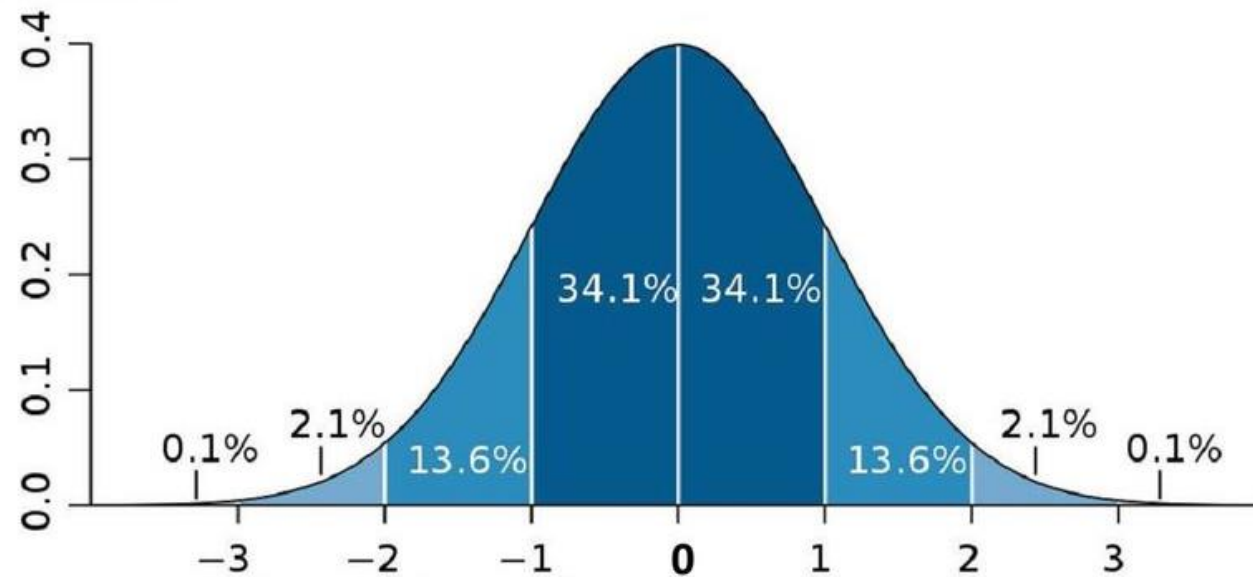
Amplitude interquartil



0.00087



# Variabilidade: Desvio-padrão



```
stddevpop(/net.tcp.service.perf[http,"{HOST.CONN}","{$NGINX.STUB_STATUS.PORT}"],1h:now/h)
```

```
stddevsamp(/net.tcp.service.perf[http,"{HOST.CONN}","{$NGINX.STUB_STATUS.PORT}"],1h:now/h)
```

# Variabilidade: Desvio Absoluto Mediano da Mediana

```
mad(//net.tcp.service.perf[http,"{HOST.CONN}","{$NGINX.STUB_STATUS.PORT}],1h:now/h)
```

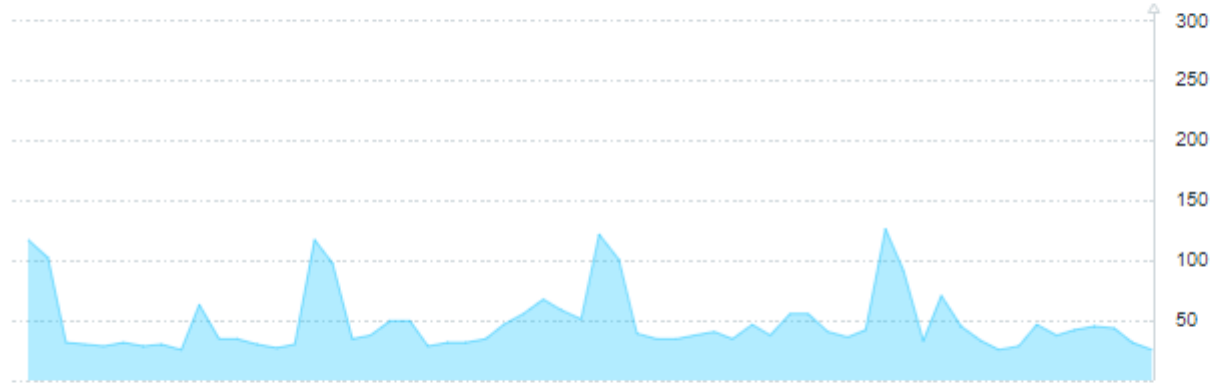


Assim como a **MEDIANA**, o **MAD não é influenciado por outliers**, sendo então uma **estimativa robusta** para cálculo de desvios.

## VARIABILITY

### ABOUT WEB SERVER REQUESTS (P/MINUTE)

WEB SERVER REQUESTS (PREVIOUS HOUR)



IQR - PREVIOUS HOUR

**20**

IQR - PREVIOUS DAY

**22**

STD DEV POP

**26**

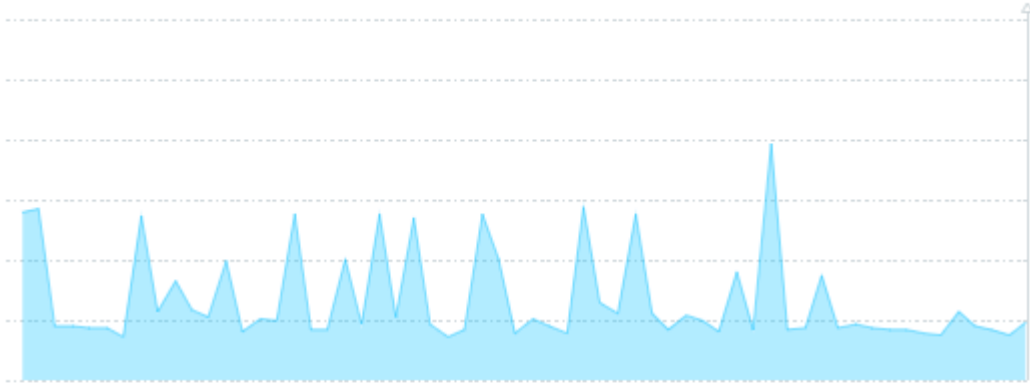
STD DEV SAMPLE

**27**

MAD

**8**

RESPONSE TIME FOR WEB SCENARIO



THE BEST

1s 734.34ms

RESPONSE TIME

THE WORST

4s 959.05ms

RESPONSE TIME

THE AVG

2s 362.55ms ↓

RESPONSE TIME

MEDIAN

1s 945.28ms

RESPONSE TIME

THE VARIANCE

0.601659 ↑

STD DEVIATION

0.78 ↑

MAD

0.14 ↑

MEDIAN ABSOLUTE DEVIATION



**3<sup>o</sup> e 4<sup>o</sup> momentos:  
Assimetria e Curtose**

# Assimetria e Curtose

- **Assimetria:** indica se um dado está inclinado a valores maiores ou menores

```
skewness(/host/key,1h)
```

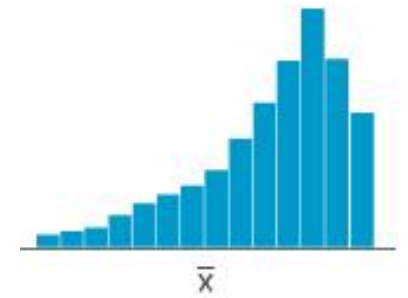
## Value mapping

\* Name

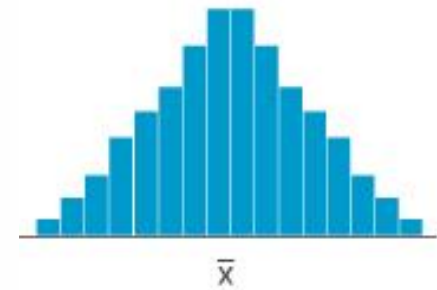
\* Mappings

Type	Value	Mapped to
is greater than or equals	1	Positive
is less than or equals	-1	Negative
equals	0	Symetric

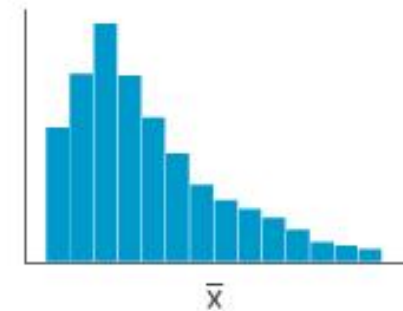
[Add](#)



Assimétrica negativa



Simétrica



Assimétrica positiva

# Assimetria e Curtose

- **Coefficiente de assimetria (Pearson)**: indica o grau de assimetria: moderada ou forte

$$\frac{((\text{last}([avg/trendavg]) - \text{last}([median])) * 3) / \text{last}([stddev]))}{}$$

## Value mapping

\* Name

\* Mappings

Type	Value	Mapped to	Action
in range	0.15-1	Moderated	<a href="#">Remove</a>
is greater than or equals	1.1	High	<a href="#">Remove</a>

[Add](#)

Update

Cancel



# Assimetria e Curtose

- **Curtose**: indica o a distribuição é propensa a picos (outliers)

```
kurtosis(/host/key,1h)
```

## Value mapping

\* Name

Kurtosis

\* Mappings

Type	Value	Mapped to
equals	0	⇒ No outliers
is greater than or equals	0.1	⇒ Prone to spikes
is less than or equals	-0.1	⇒ Not prone to spikes

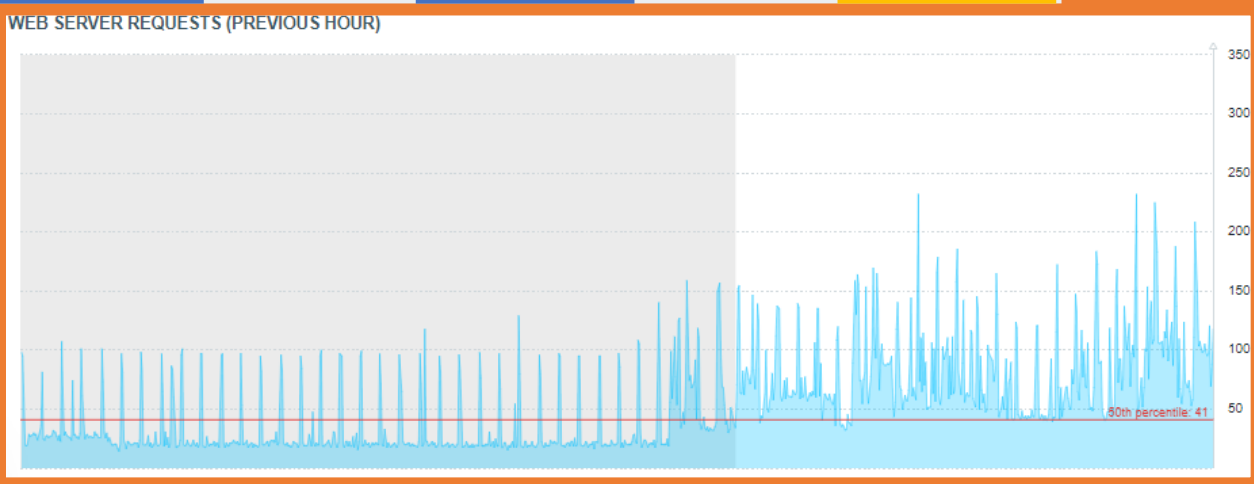
# Aplicabilidade e Insights

# Requisições à aplicação Web

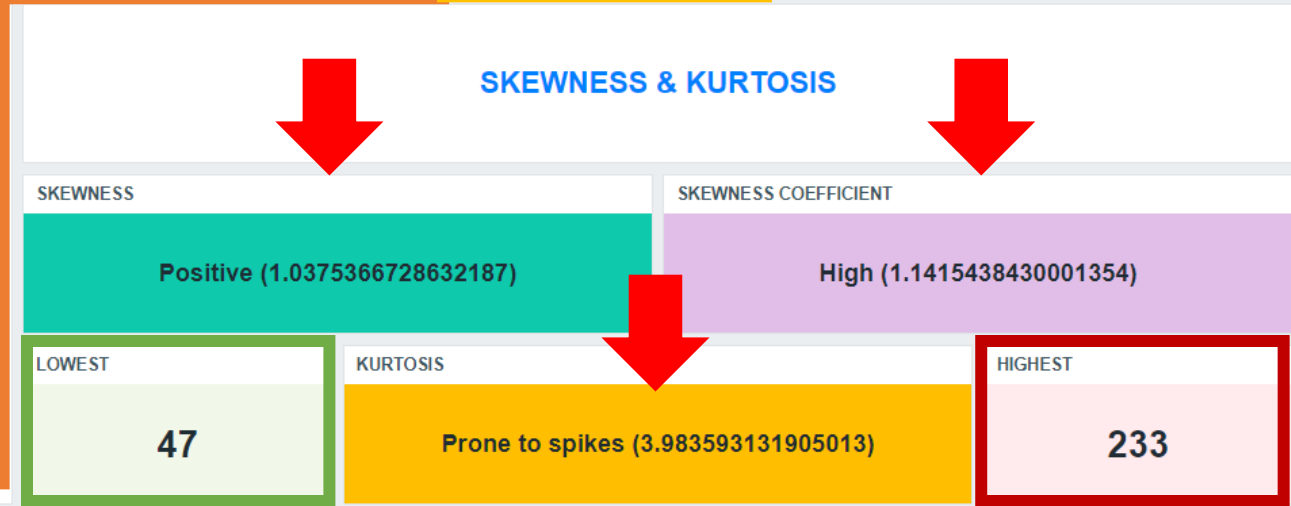
LOCATION ESTIMATE



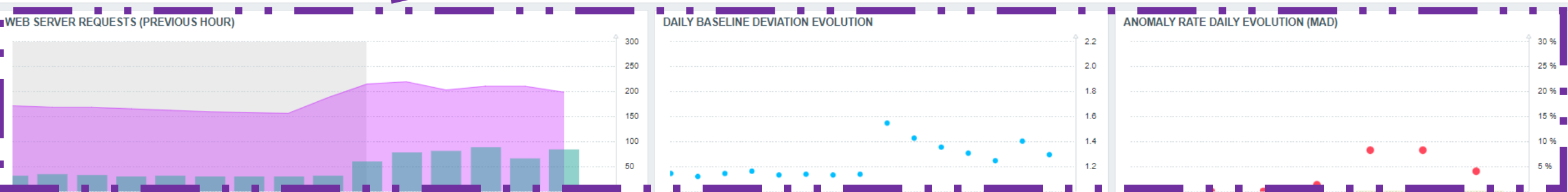
VARIABILITY



SKEWNESS & KURTOSIS



BASELINE MONITORING & ANOMALY RATE



# Ponderações

- **Acessos:** a aplicação tem **mais acessos durante o dia**.
- **Desvios e assimetria:** não há padrão nos acessos. Esperado de um serviço web? Drilldown em páginas específicas?
- **Pode haver novidade?** por que tantos acessos em determinado momento? Por que tão pouco em outros?
- **Baseline:** por que houve menos acessos à página que o esperado? **Estamos desinteressantes?**
- **Desvios e anomalias:** as anomalias detectadas sugerem **fortes mudanças nos padrões comportamentais**. O que houve?

# Tempos de resposta da aplicação Web

LOCATION ESTIMATE (BASED ON AT THE PREVIOUS HOUR)

**AVG**  
 76.48ms ↓

**MEDIAN**  
 71.1ms ↑

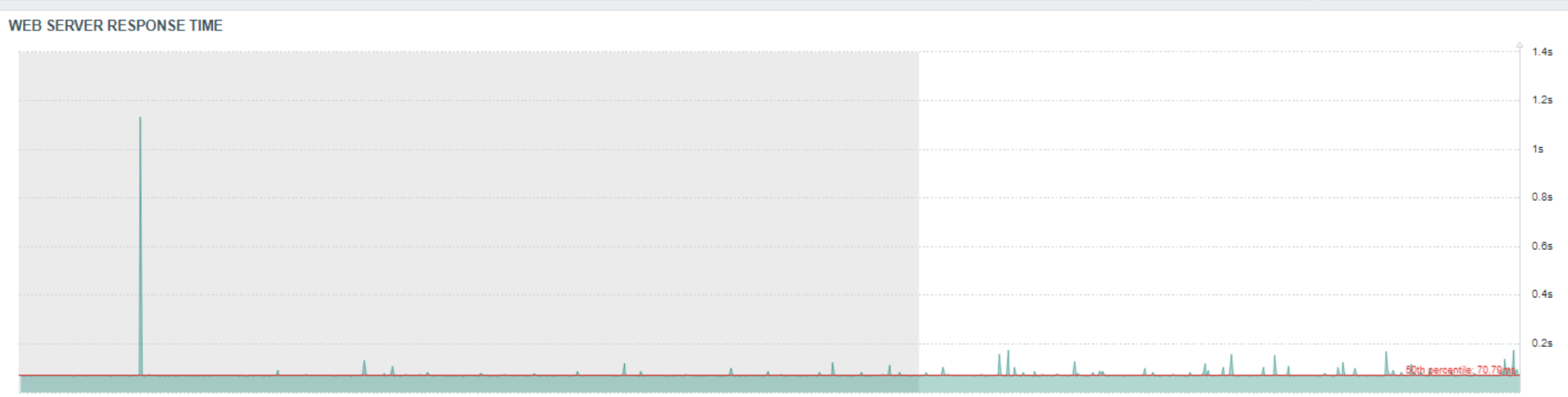
25° PERCENTILE: 70.66ms ↑

50° PERCENTILE: 71.09ms ↑

75° PERCENTILE: 71.54ms ↓

**THE BEST**  
 69.58ms ↑

**THE WORST**  
 175.35ms ↑



**VARIABILITY**

BEST < WORST PREVIOUS HOUR: 60.32 % ↑

BEST < WORST TODAY SO FAR: 93.88 %

IQR PREVIOUS HOUR: 0.88ms ↓

IQR TODAY SO FAR: 0.8ms ↑

DEVIATION POPULATION: 0.02 ↑

DEVIATION SAMPLE: 0.02 ↑

DEVIATION MAD: 0.00 ↓

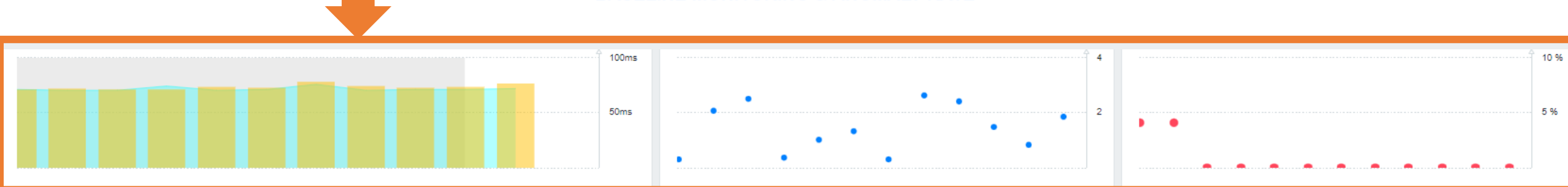
**SKEWNESS & KURTOSIS**

SKEWNESS: Positive (4.183830771839158) †

SKEWNESS COEFFICIENT: Moderated (0.9441991409790612) †

KURTOSIS: Prone to spikes (21.842779253010477) †

BASELINE MONITORING & ANOMALY RATE





# Ponderações

- **Performance:** Acessos não padronizados e tempos de resposta adequados: Performance OK da aplicação.
- **Desvios:** Poucos desvios, tempos de resposta adequados e dentro da média.
- **Assimetria:** Moderada, confirmando a baixa variação dos tempos de resposta.
- **Baseline:** Por que houveram menos acessos à página que o esperado?
- **Desvios e Anomalias:** Sem anomalias, o que reforça que os desvios não representam alterações significativas de performance no ambiente.

**Muito ainda há se explorar!**



 [instagram.com/uniredeinfo/](https://www.instagram.com/uniredeinfo/)

 [linkedin.com/company/unirede/](https://www.linkedin.com/company/unirede/)

 [facebook.com/unirede.net](https://www.facebook.com/unirede.net)

 [youtube.com/uniredeinfo](https://www.youtube.com/uniredeinfo)

 [twitter.com/unirede](https://twitter.com/unirede)

Obrigado!