

# Valores Fora do Padrão: Outliers

- Podem causar distorções em relatórios
  - Uma venda atípica faz pensar que todo o mês terá o mesmo padrão de vendas
  - Uma venda atípica pode fazer previsões acima ou abaixo do que seria mais provável



# Modelo de Previsão de Vendas

- A partir de investimento em anúncio em mídias sociais, queremos prever quanto vamos vender

Investimento	Vendas
51,72	12,12
273,6	25,8
243	19,92
212,4	20,52
352,32	24,84
248,28	15,48
30,12	10,2
210,12	17,88
107,64	12,72
287,88	27,84
272,64	17,76
80,28	11,64
239,76	13,68
120,48	12,84
259,68	27,12
219,12	25,44
315,24	24,24
238,68	28,44
8,76	6,6
163,44	15,84
252,96	28,56
252,84	22,08
64,2	9,72
313,56	29,04
287,16	18,84
123,24	16,8
157,32	21,6
82,8	11,16
37,8	11,4

Alboukadel Kassambara (2019). datarium: Data Bank for Statistical Analysis and Visualization. R package version 0.1.0. <https://CRAN.R-project.org/package=datarium>

# Modelo de Previsão de Vendas



- Quanto vou vender se investir 180?
  - Saída do modelo: 18

# Outlier nos dados

- Quanto vou vender se investir 180:
  - Saída do modelo: 23

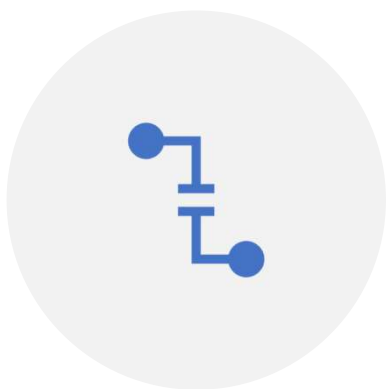
Investimento	Vendas
51,72	12,12
273,6	25,8
243	19,92
212,4	20,52
352,32	24,84
248,28	15,48
30,12	10,2
210,12	17,88
107,64	12,72
287,88	27,84
272,64	200
80,28	11,64
239,76	13,68
120,48	12,84
259,68	27,12
219,12	25,44
315,24	24,24
238,68	28,44
8,76	6,6
163,44	15,84
252,96	28,56
252,84	22,08
64,2	9,72
313,56	29,04
287,16	18,84
123,24	16,8
157,32	21,6
82,8	11,16
37,8	11,4

## "Conclusão"

- Outliers podem ser um problema para a análise de dados, não podemos simplesmente "ignorá-los"
- Porém não existe regras gerais ou processos definidos para seu tratamento



# Avaliação



O VALOR É UM ERRO DE  
ENTRADA?



O VALOR É UMA  
VARIÁVEL DE INTERESSE?



O VALOR É UMA  
EXCEÇÃO?

# Erro de Entrada



Digitação errada



Sensor de equipamento mal calibrado



Erro de Sistema



Erro de processo de carga de dados

O valor é uma  
variável de  
interesse?

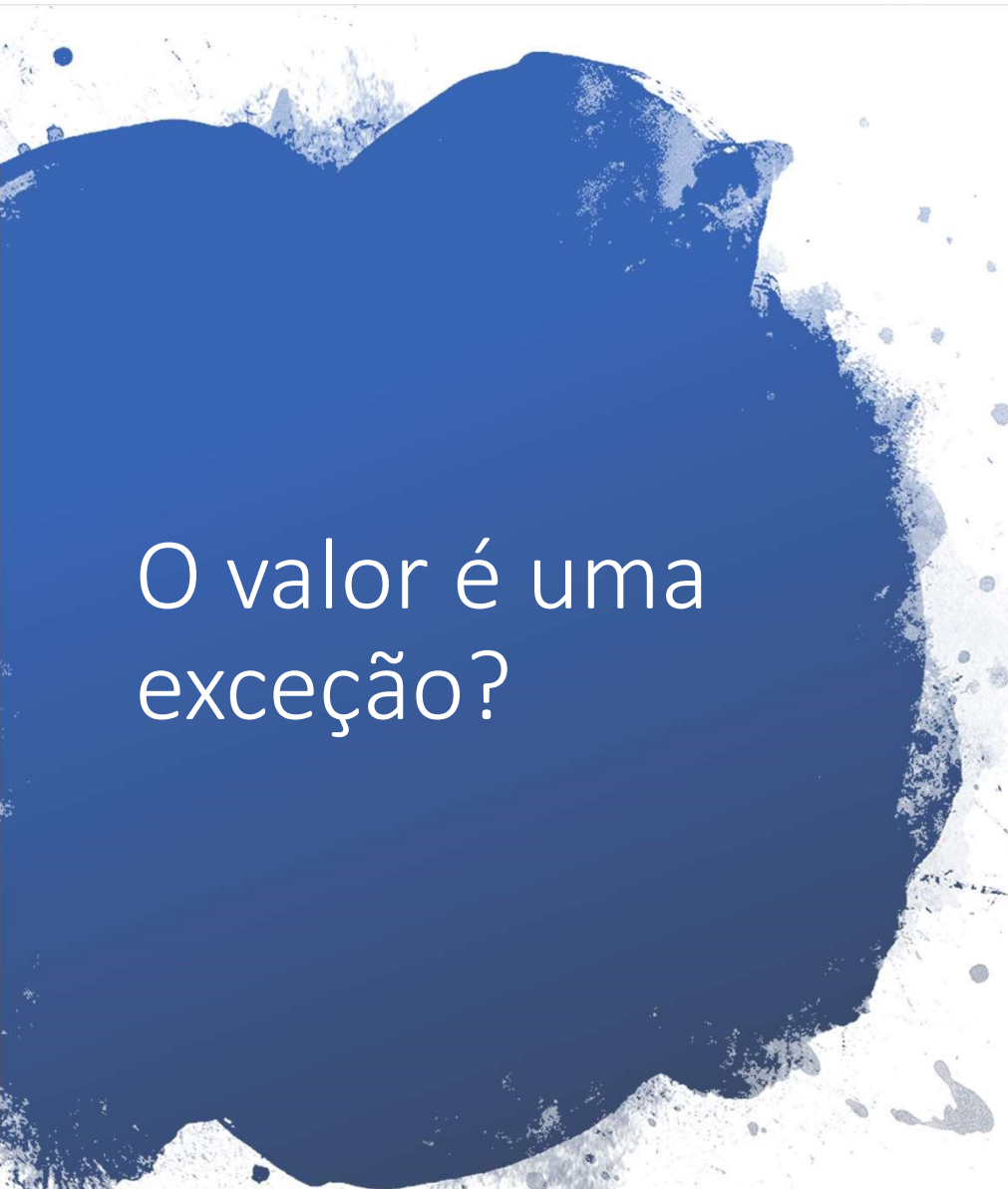


Um produto que não esta no escopo



Um salário não "elegível"





O valor é uma  
exceção?

- Uma venda excepcional
- Uma promoção
- Um feriado prolongado



# Como identifica-los?

- Distância da média:
    - 2 ou 3 desvios padrão da média
-



O que fazer? Tratamento

**Remover**

**Substituir  
pela Mediana**

# Porém...

- Você não deve tratar outliers que não são "erro de entrada" quando:
  - Produção de relatórios de conferência
  - Gráficos / painéis
- Você pode tratar:
  - Projeções
  - Produção de modelos preditivos
  - Cálculos de parâmetros estatísticos quando o dados não faz parte do escopo



**PROF.  
FERNANDO  
AMARAL**  
[www.datascientist.com.br](http://www.datascientist.com.br)