

UNILEÃO
Centro Universitário

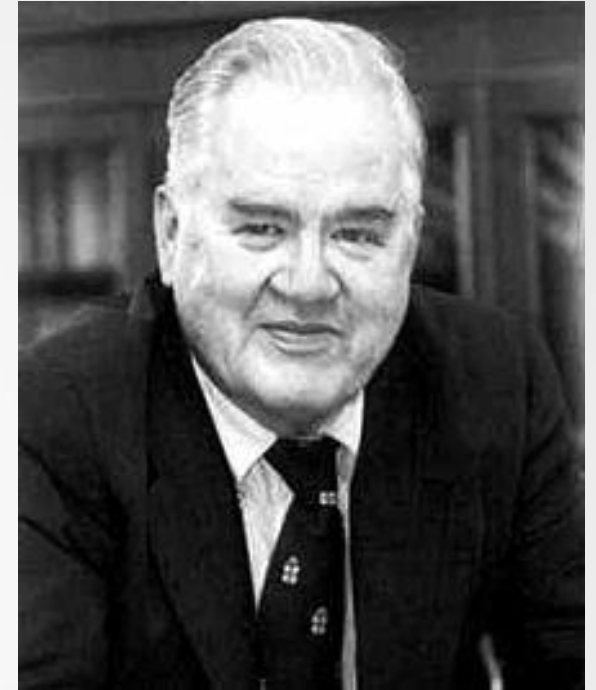
Mineração de dados

Análise exploratória

Curso de Análise e Desenvolvimento de Sistemas

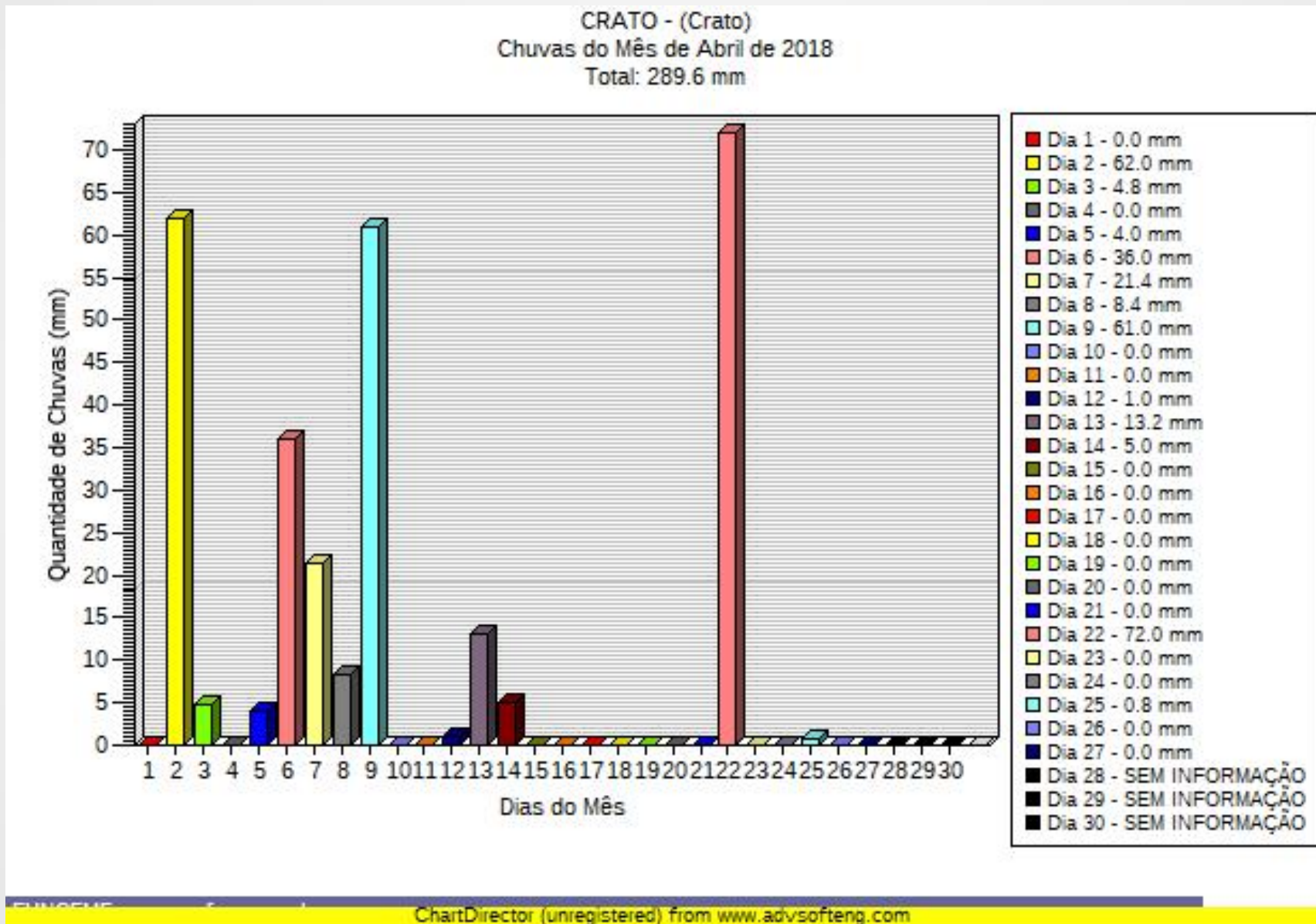
- **John Tukey:**

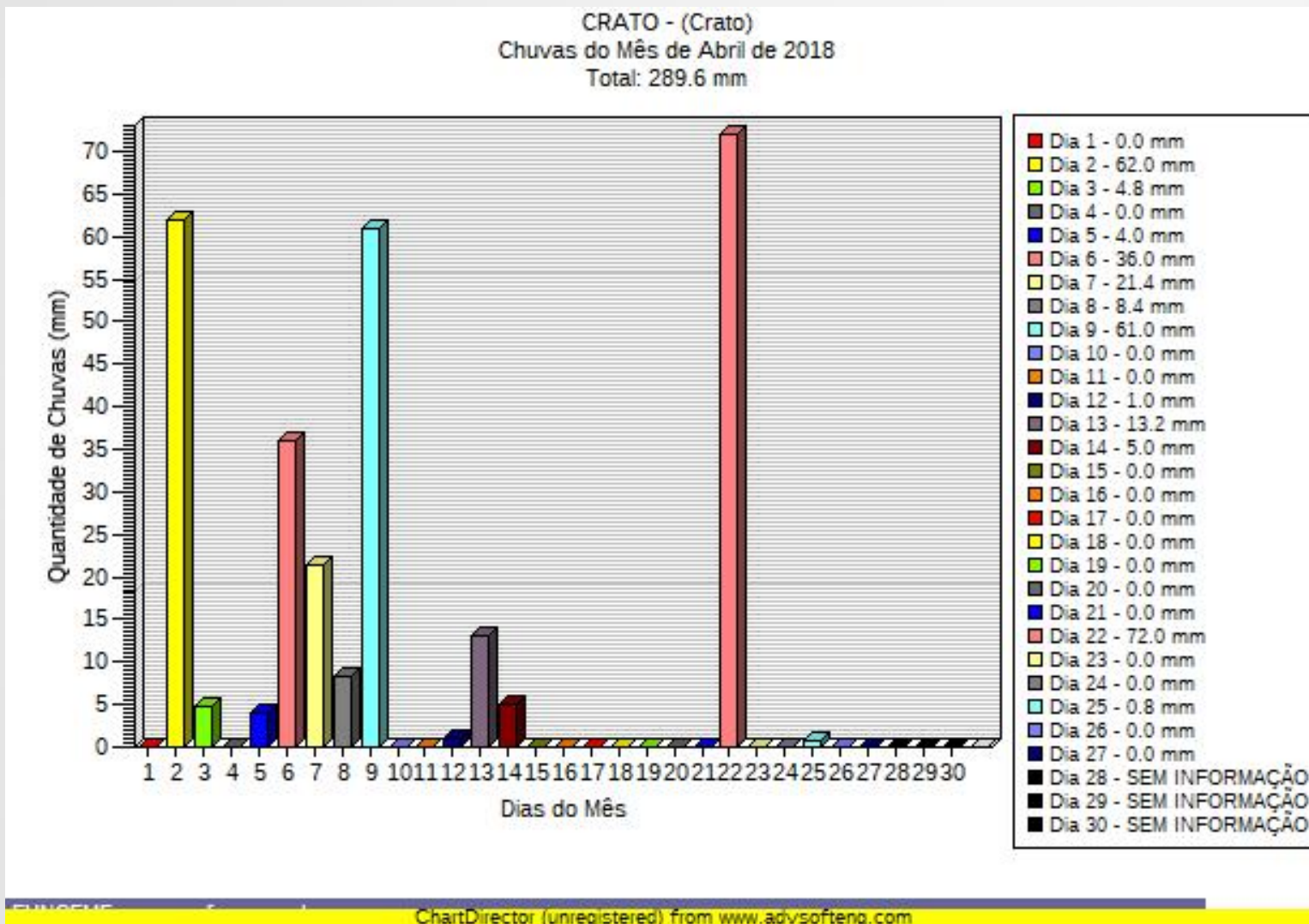
- EUA;
- Estatístico;
- Livros:
 - *Exploratory Data Analysis* (1977);
 - *Understanding Robust and Exploratory Data Analysis* (1983);
 - *Exploring Data Tables, Trends and Shapes* (1985);
- Criador do termo **bit**.



- Estatística;
- Visualização;
- *Online Analytical Processing (OLAP).*

- Frequência;
- Moda;
- Média;
- Faixa;
- Mediana;
- Variância.





$$f_i = \frac{n_i}{N} = \frac{n_i}{\sum_i n_i}$$

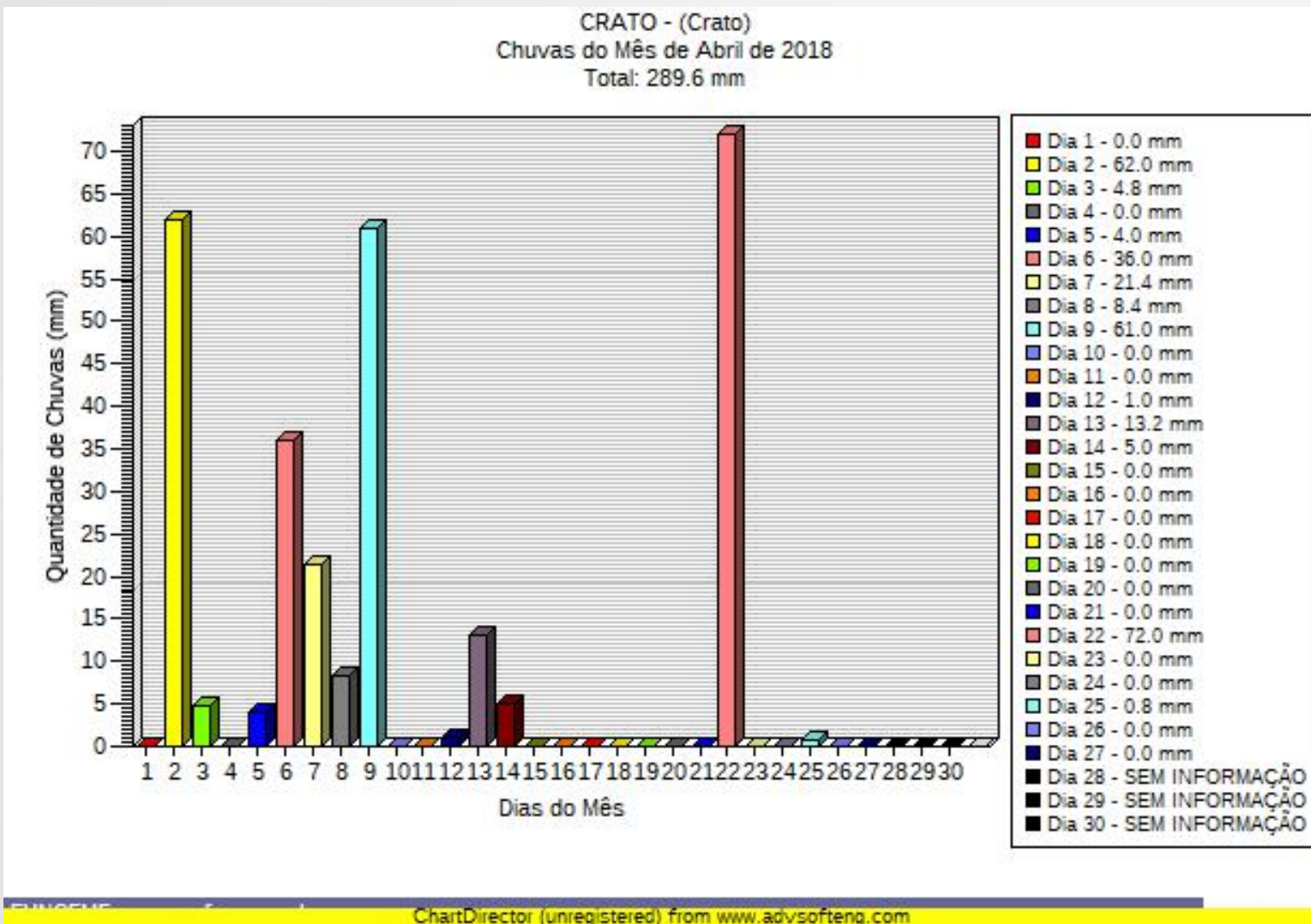
- Ex: Frequência de não-ocorrências de chuva

Dias s/ chuva (n_i) = 15

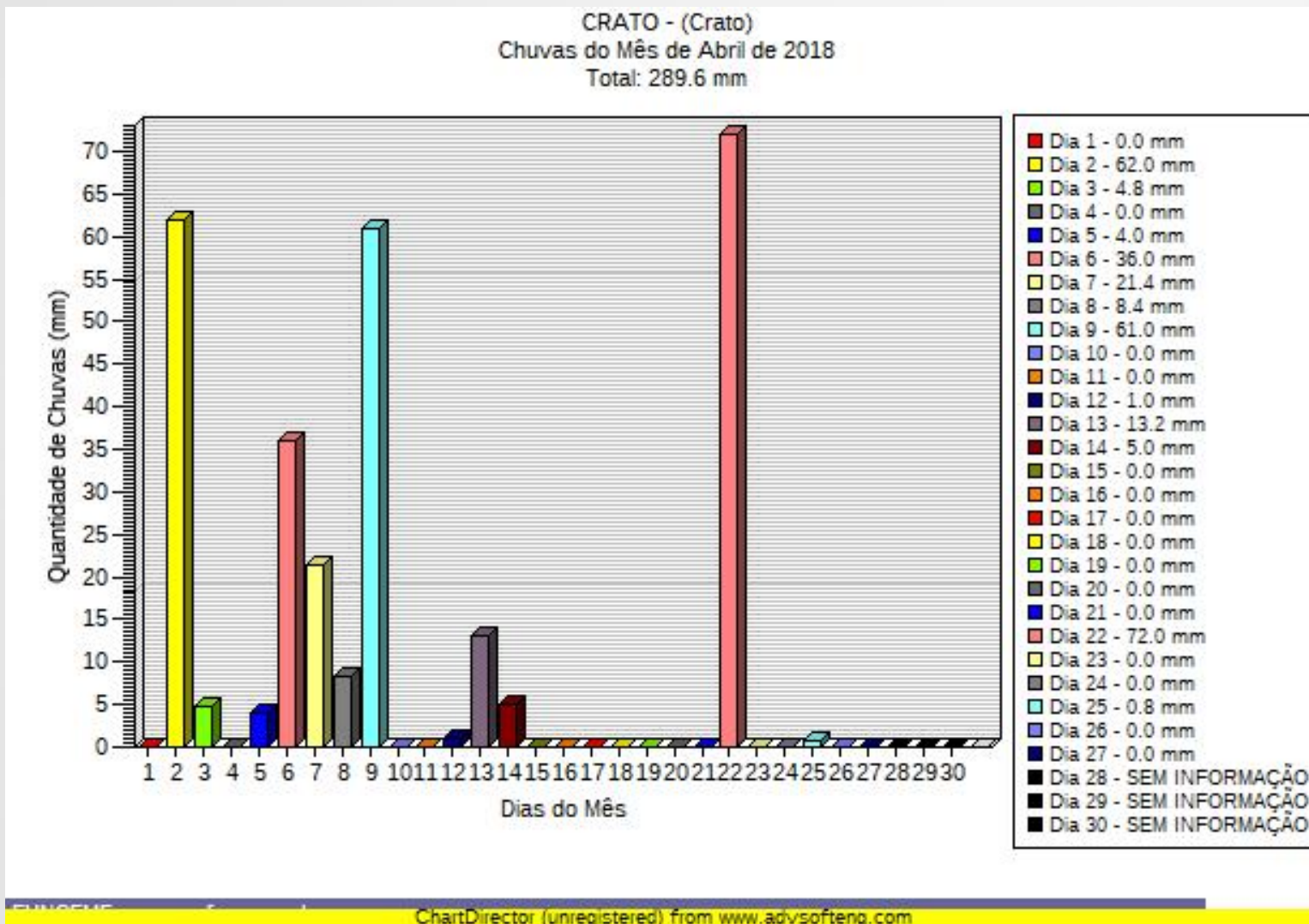
Total de dias (N) = 27

$$f_i = 15/27$$

$$f_i = 0,5555... = 55,56\%$$



- Evento que mais se repete. Neste caso: 0mm



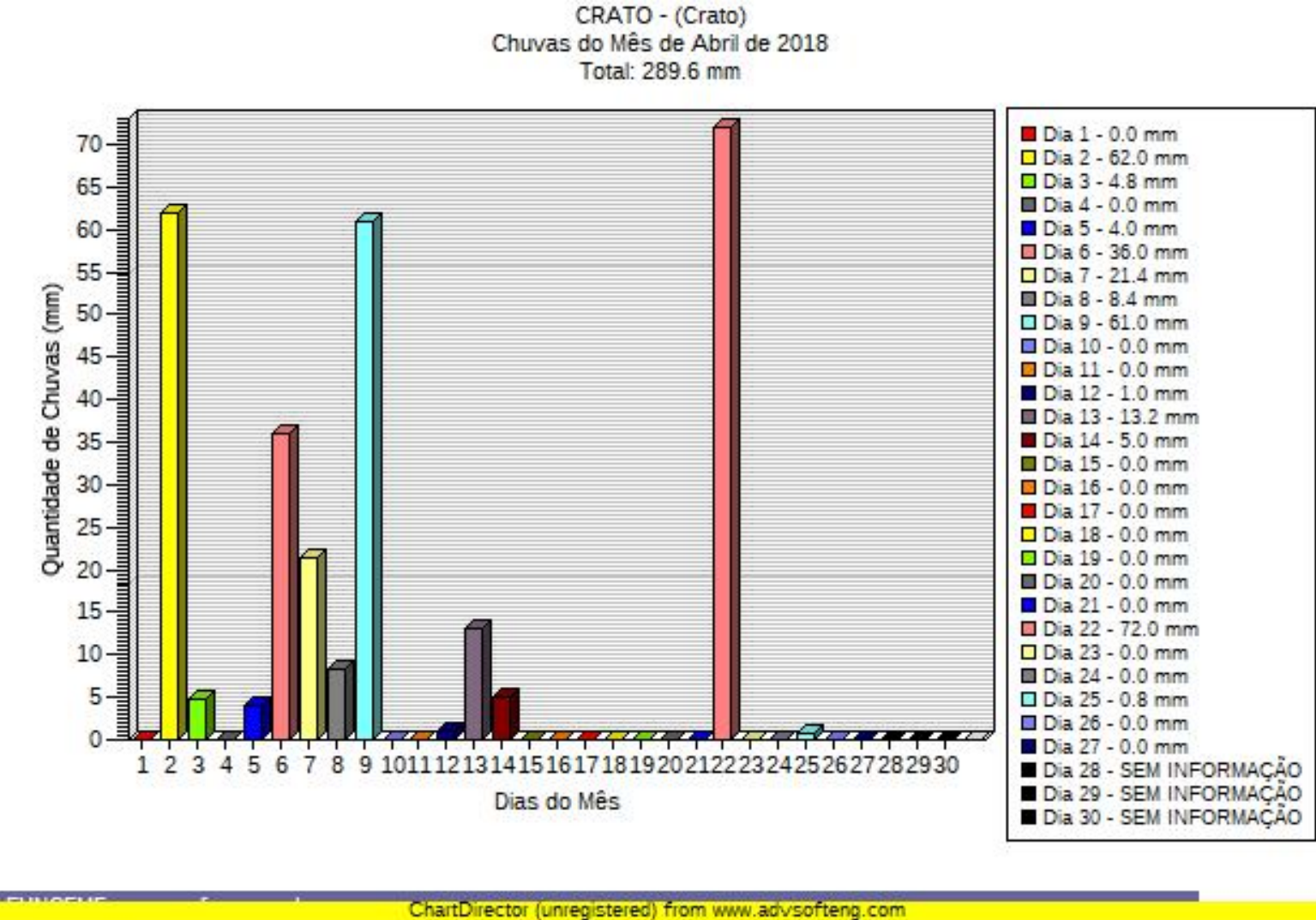
$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$

- Soma dos valores (Σ) de todos os eventos dividido pela quantidade (n) de eventos.

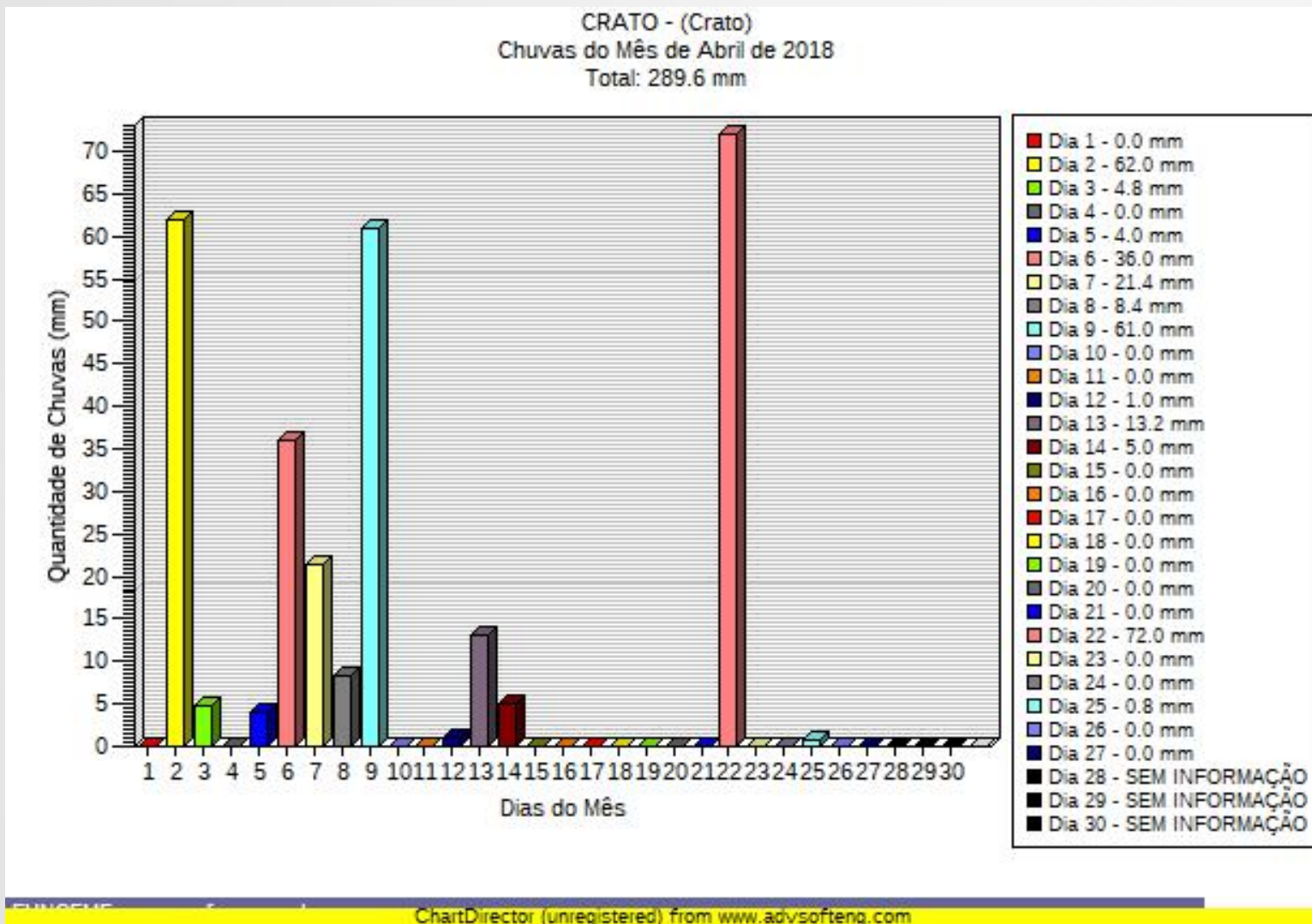
- $\Sigma = 289,6$

- $n = 27$

$$\bar{x} = 10,73\text{mm}$$



- A faixa é delimitada pelo maior e menor valor. Já a mediana é o que está exatamente no meio.
- Faixa: 0mm a 72mm
- Mediana: 36mm



- A variância é a medida de quanto cada amostra se distancia da média.