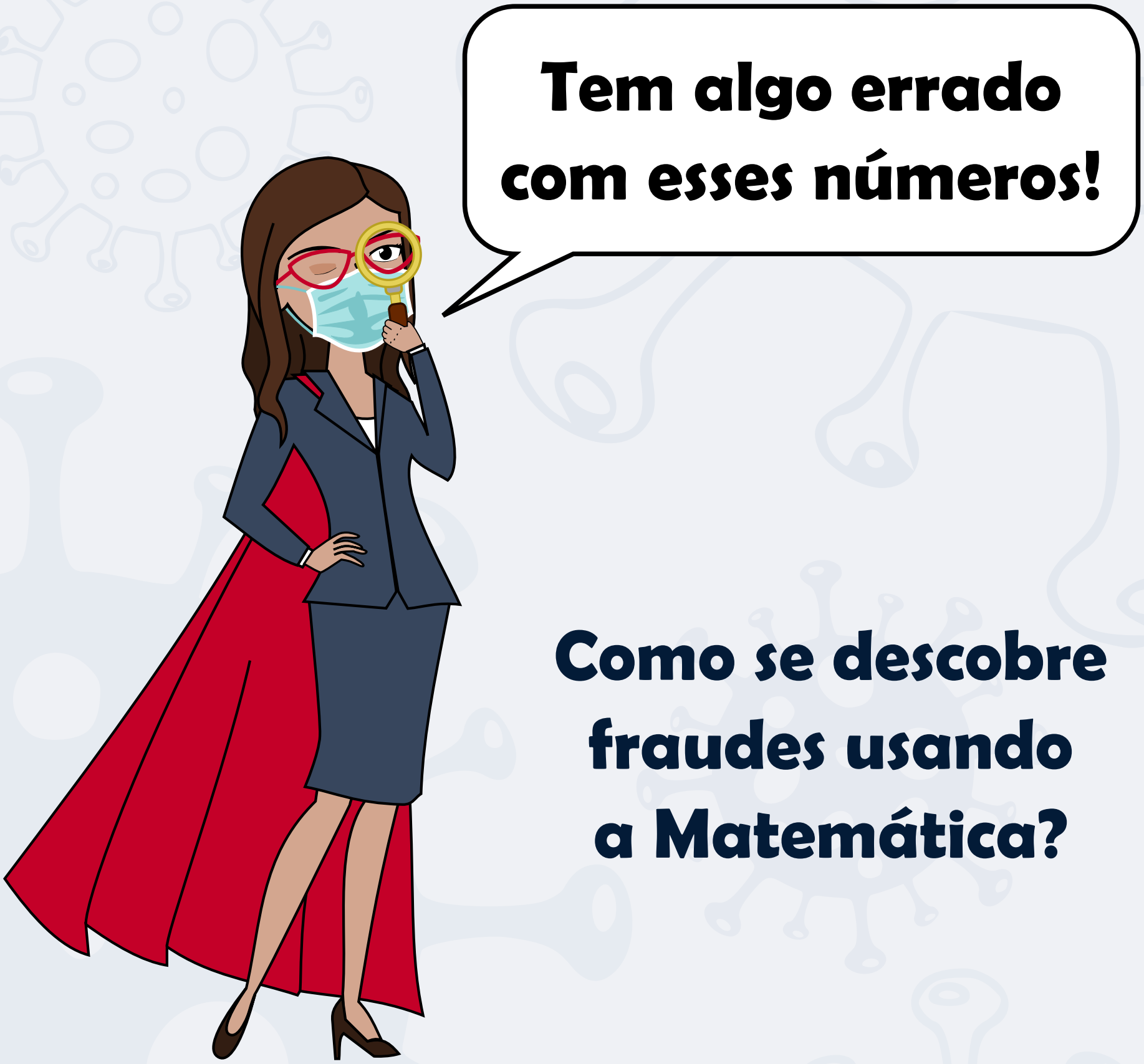
Matemática das Epidemias



Como se descobre fraudes usando a Matemática?

uff UFSC unipampa

Repare uma coisa...

Se considerarmos, por exemplo, as alturas das 58 construções mais altas no mundo, um padrão interessante aparece. Para 24 delas, o primeiro dígito da altura (em metros) é 1, enquanto para apenas uma delas o primeiro dígito é 9:

1º dígito	Nº de construções	% do total
1	24	41,4%
2	9	15,5%
3	7	12,1%
4	6	10,3%
5	1	1,7%
6	5	8,6%
7	1	1,7%
8	4	6,9%
9	1	1,7%

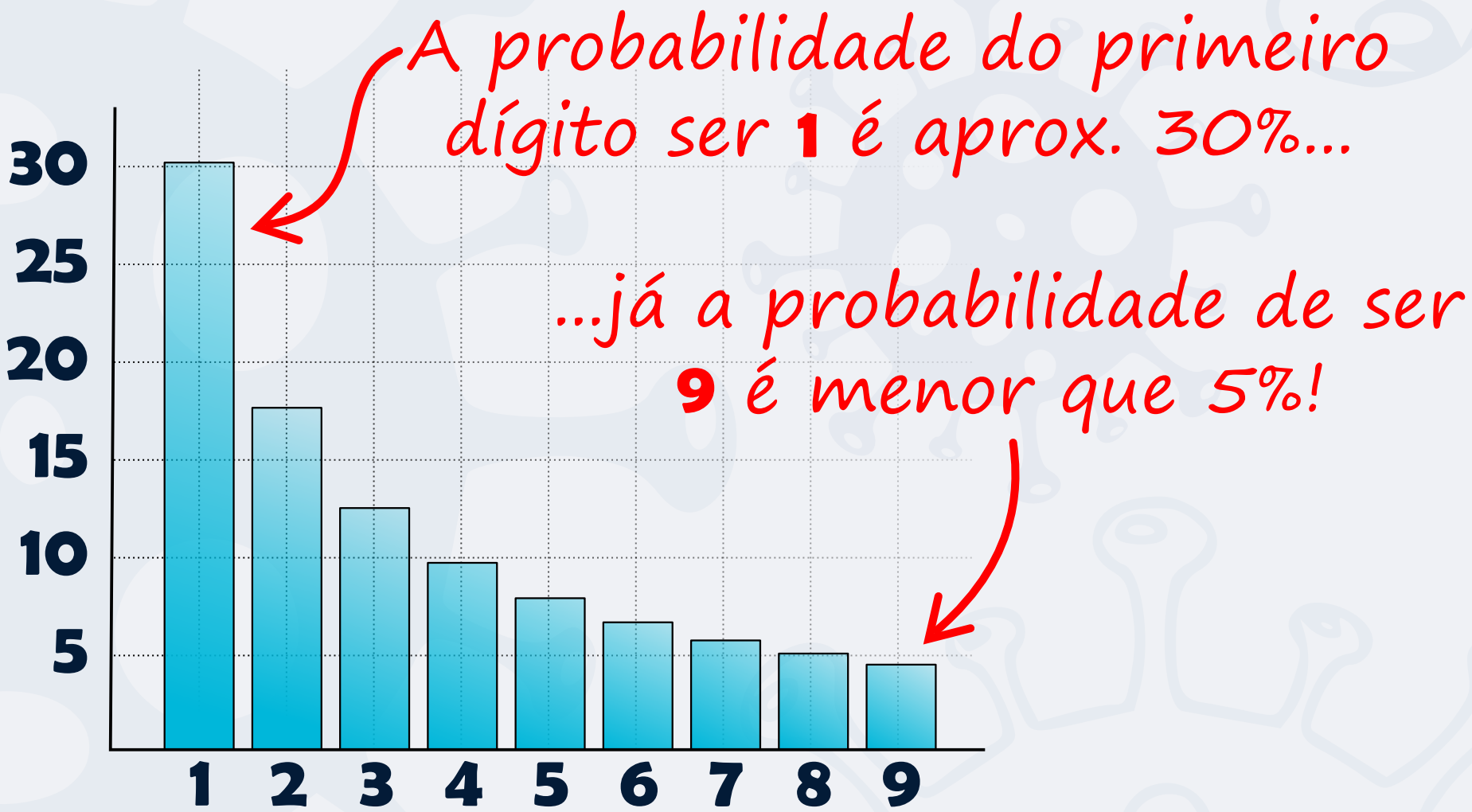
Tende a decrescer...

uff UFSC unipampa

A lei de Benford

Esse exemplo é uma manifestação da Lei de Benford:

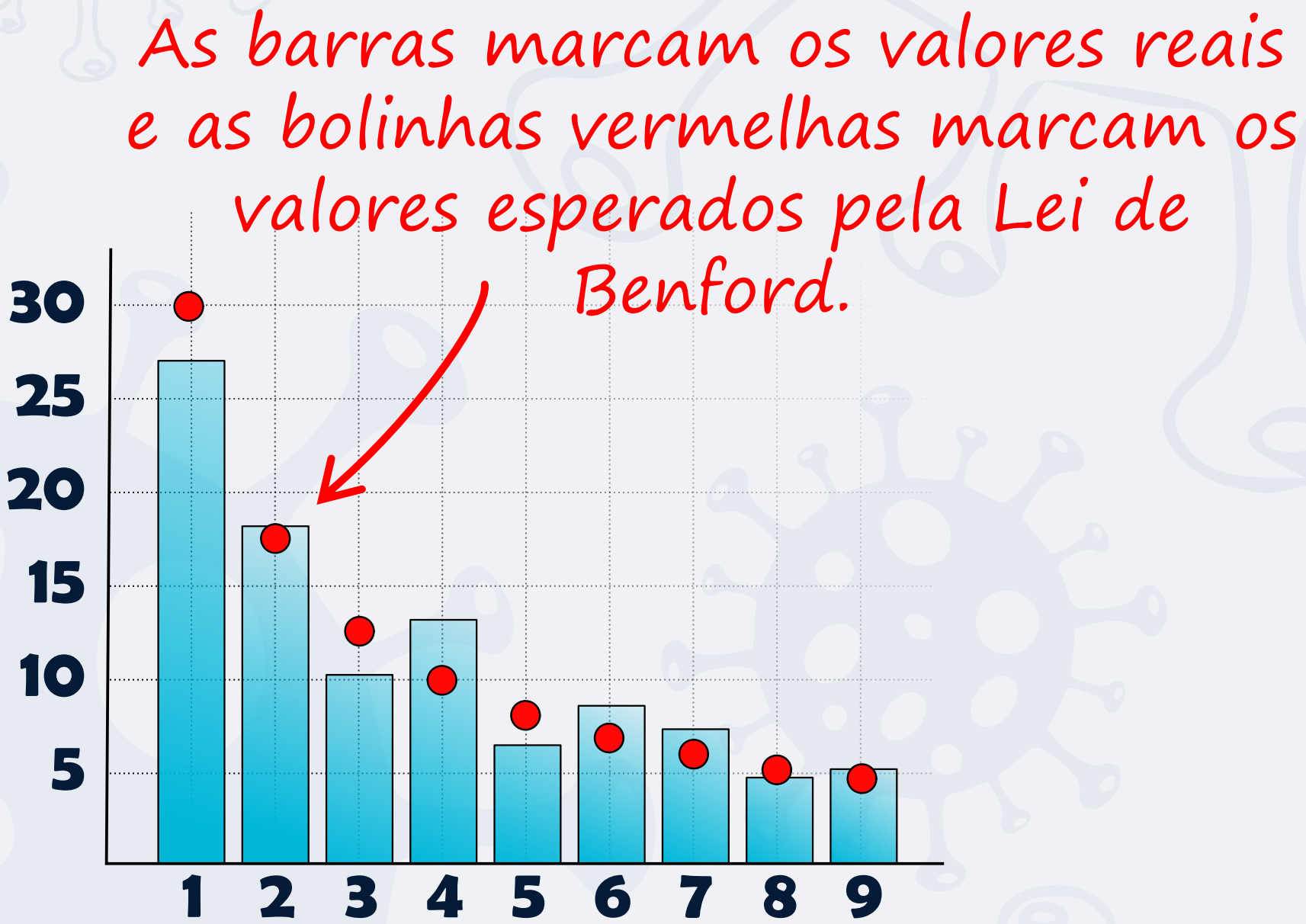
Num conjunto de dados numéricos que abrange uma ampla gama de valores, é mais provável que o primeiro dígito de um número aleatório seja pequeno:



uff UFSC unipampa

Exemplificando...

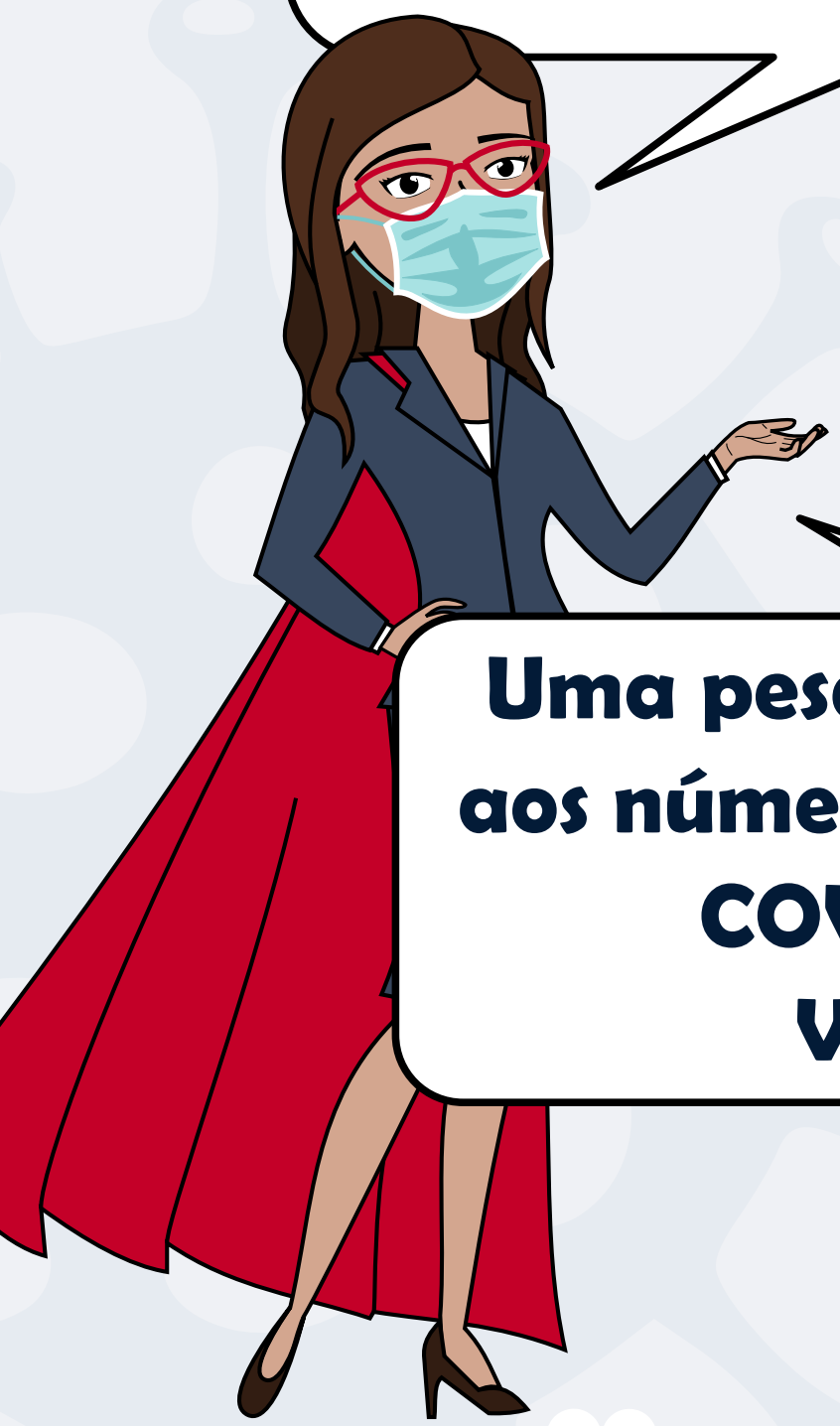
Outro exemplo: abaixo vemos a distribuição dos primeiros dígitos das populações dos 237 países do mundo (em 2010):



Quando há fraude nos dados, o fraudador geralmente não observa a Lei de Benford, e os valores diferem muito do esperado.

Aplicações

A Lei de Benford é muito usada, portanto, para se detectar fraudes em finanças, declarações de impostos, dados eleitorais, dados científicos...
...e, inclusive, dados sobre a COVID-19!

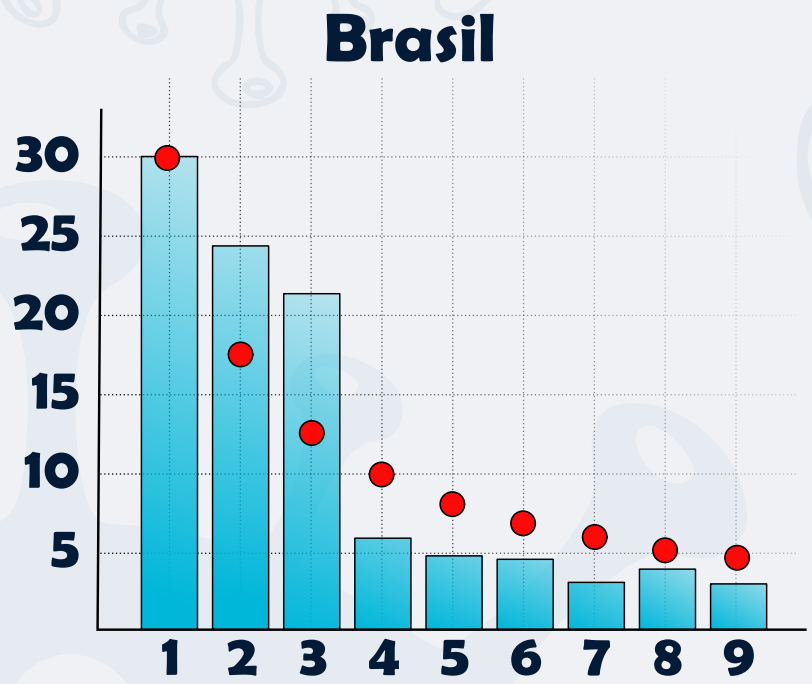


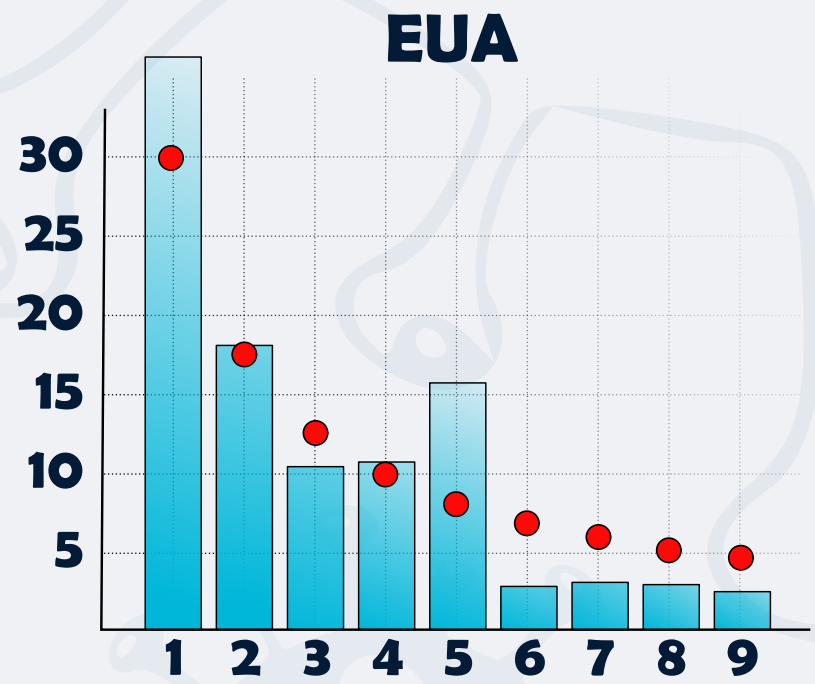
Uma pesquisa recente[1] aplicou essa lei aos números de infectados e mortos por COVID-19 em vários países.
Vejamos os resultados.

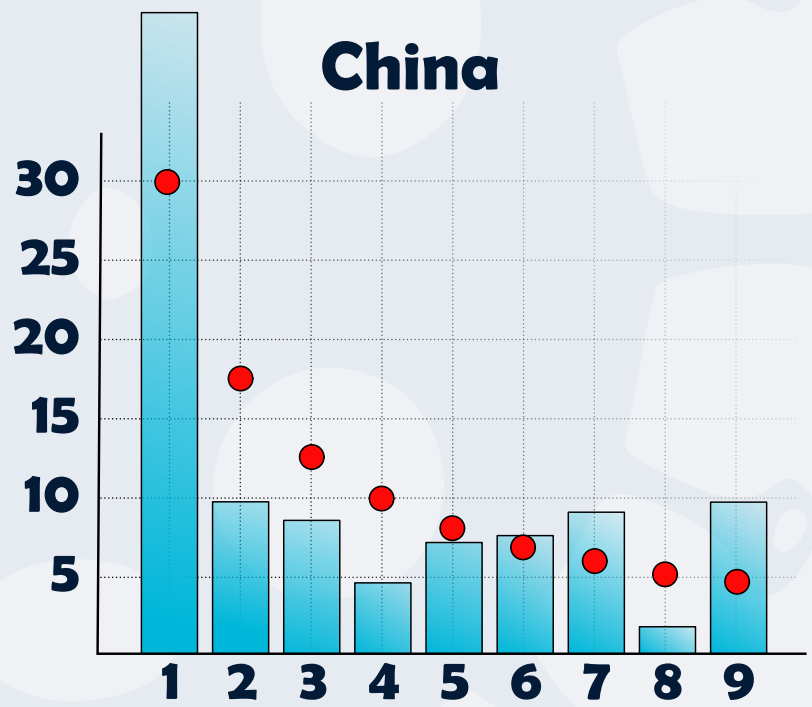
[1]https://cutt.ly/6hsRKd2

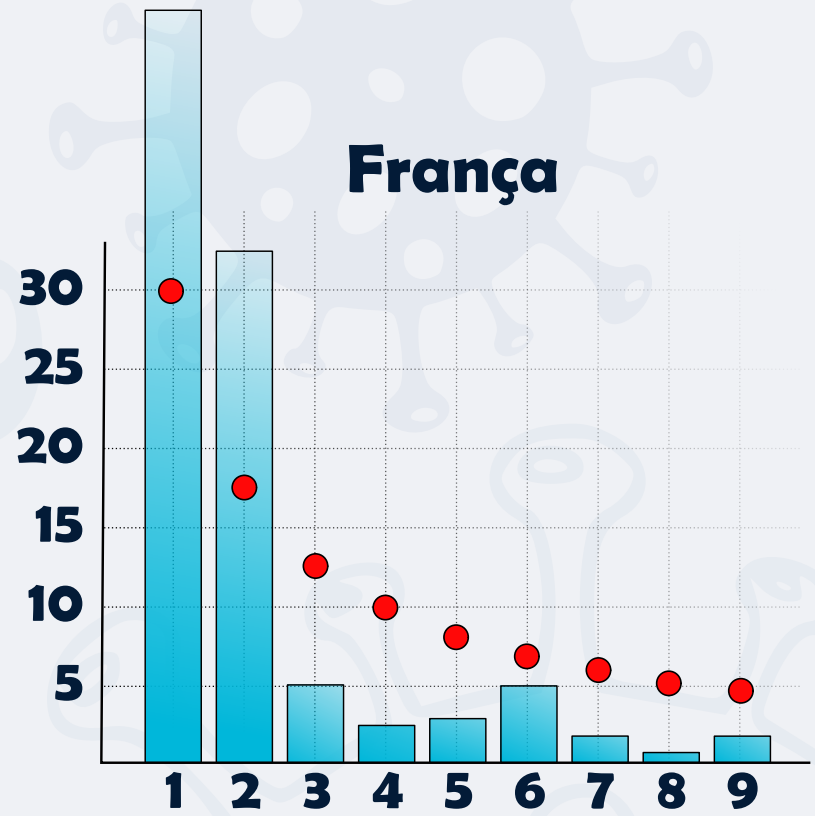
Total de casos

Não foram encontradas divergências significativas nos dados da maioria dos países, dentre eles o Brasil:





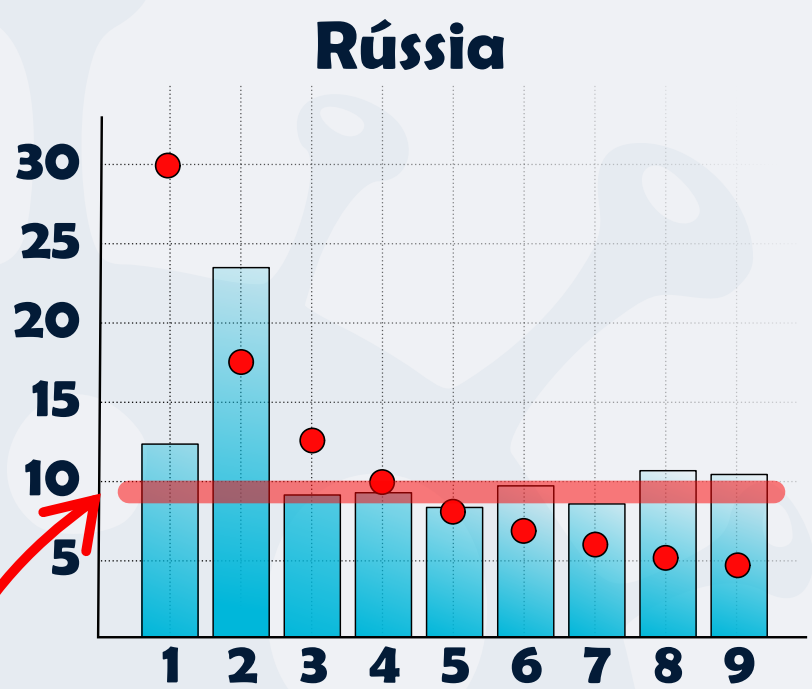


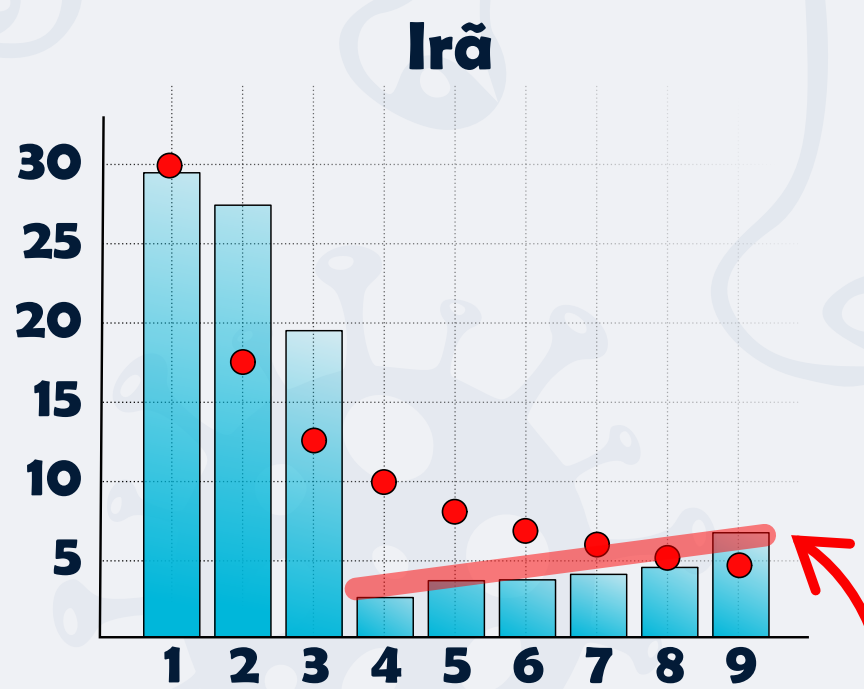


uff UFSC unipampa

Total de casos

A análise, no entanto, indicou suspeitas de manipulação nos dados da Rússia e do Irã:





O fato dos dados da Rússia serem quase constantes, e do Irã crescerem do 4 em diante, sugerem fraude.

Para os curiosos

A Lei de Benford se expressa mais precisamente com logaritmos, portanto é uma aplicação interessante deste objeto:

A probabilidade do primeiro dígito de um número (num conjunto que satisfaz a lei) ser d é

$$\mathrm{P}(d) = \log_{10}\left(1 + \frac{1}{d}\right)$$

Outra curiosidade é que a lei não depende das unidades de medida utilizadas. Além disso os números não precisam estar escritos na base 10. Para uma base b qualquer, tem-se:

$$\mathrm{P}(d) = \log_{b}\left(1 + \frac{1}{d}\right)$$

Continue ligad@!

@matematicadasepidemias

Matemática das Epidemias

cutt.ly/matematicadasepidemias

Nos próximos capítulos veremos:
- Eficiência das máscaras
- Outros modelos

uff UFSC unipampa