

A series of overlapping geometric shapes, primarily triangles and quadrilaterals, in shades of teal and purple, located in the top-left corner of the slide.

AceLeraDev Data Science

Continuação Análise de dados exploratória

A series of overlapping geometric shapes, primarily triangles and quadrilaterals, in shades of teal and purple, located in the bottom-right corner of the slide.

Revisão Probabilidades

/

- Variam de 0(impossível) a 1 (certo);
- Probabilidades somam 1;
- A distribuição relaciona x com a probabilidade $p(x)$;
- Podem ser funções discretas ou contínuas;

Revisão Probabilidades

/

- Ex: Qual a probabilidade de obter cara jogando uma moeda(justa) para cima? (Discreta)
- Ex: Qual probabilidade de uma pessoa tirar nota acima da média no Enem? (Contínua)



AceLeraDev Data Science

Continuação Análise de dados exploratória



A distribuição Gaussiana (Normal)

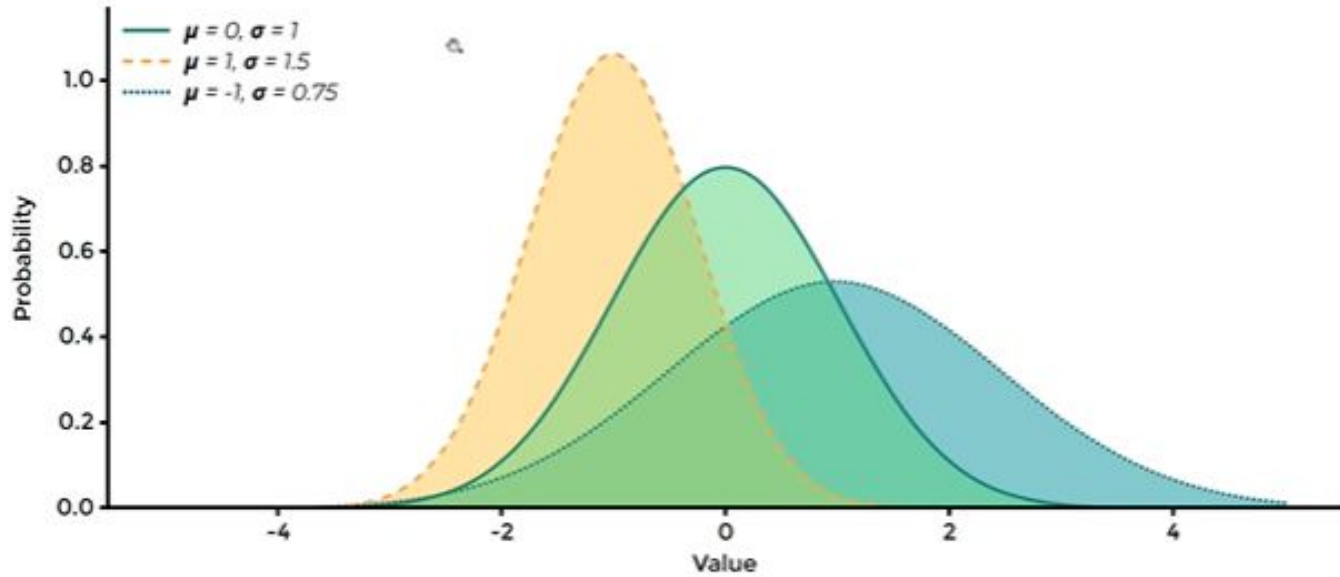
- Área sob essa curva determina a probabilidade de ocorrer o evento por ela correlacionado;
- Soma da área sob a curva de densidade igual a 1;
- Uma curva simétrica em torno do seu ponto médio, apresentando assim seu famoso formato de sino.
- Média, mediana e moda dos dados possuem o mesmo valor.

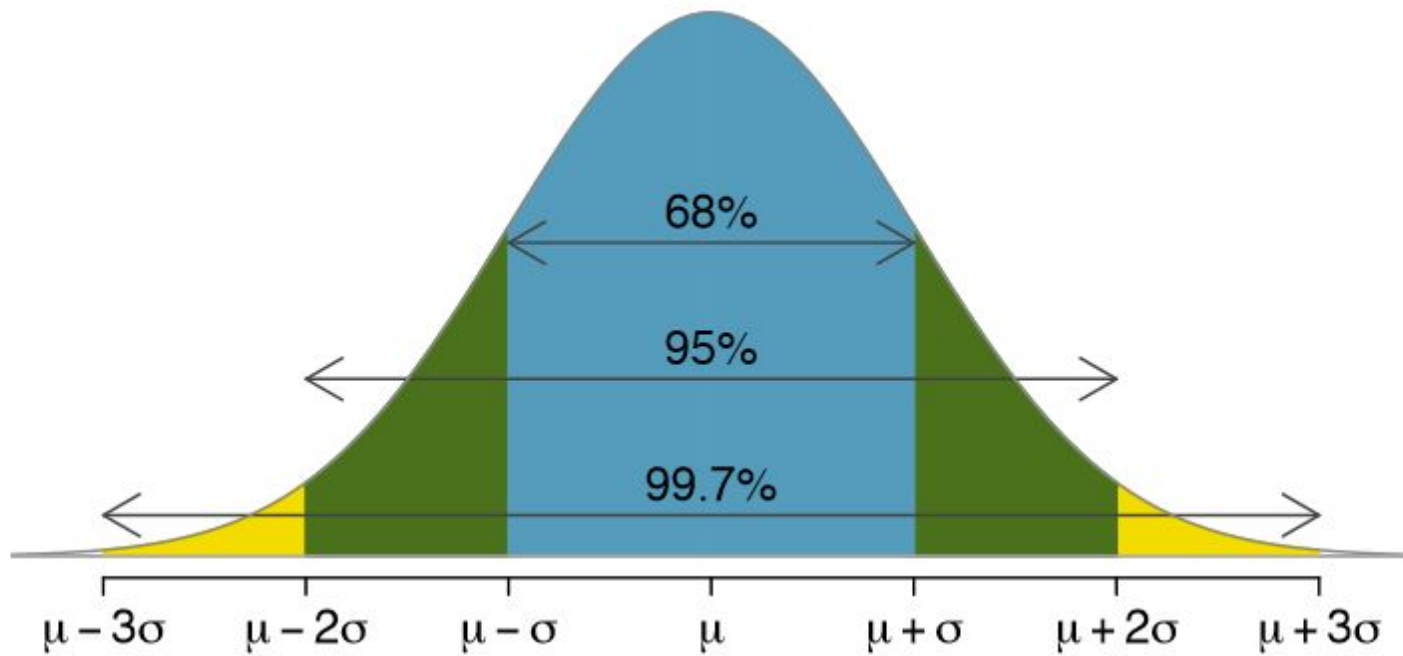


A distribuição Gaussiana (Normal)

- Altura e peso de uma população;
- Tamanho do crânio de recém nascidos;
- Pressão sanguínea;

$$p(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x - \mu)^2}{2\sigma^2}\right]$$





The diagram illustrates the formula for calculating a z-score. The formula is
$$z = \frac{x_i - \mu}{\sigma}$$
 where z is the z-score, x_i is the data point, μ is the mean, and σ is the standard deviation. Arrows point from the text labels to the corresponding parts of the formula: 'The z-score' points to z , 'The data point in question' points to x_i , 'The mean' points to μ , and 'The standard deviation' points to σ .

The z-score

The data point in question

The mean

The standard deviation

$$z = \frac{x_i - \mu}{\sigma}$$



AceLeraDev Data Science

Continuação Análise de dados exploratória



A distribuição Binomial

/

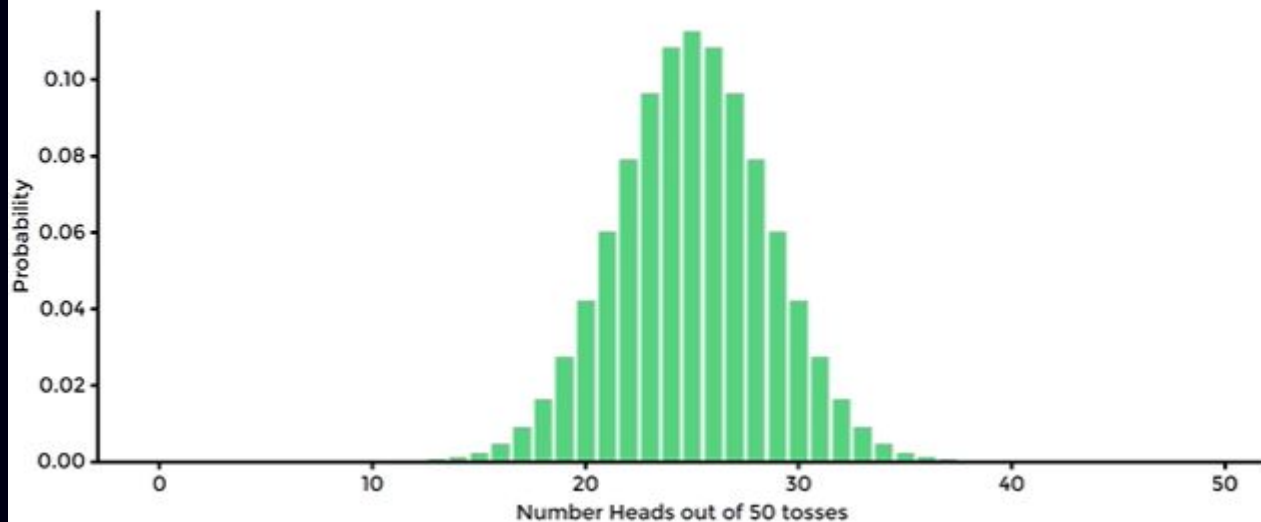
- Distribuição de probabilidade e estatística discreta (inteiros e contáveis) uma determinada sequência de tentativas;
- Espaço amostral finito;
- Apenas dois resultados possíveis tentativa;
- Todos os elementos devem possuir possibilidades iguais de ocorrência;
- Eventos devem ser independentes um dos outros;

A distribuição Binomial

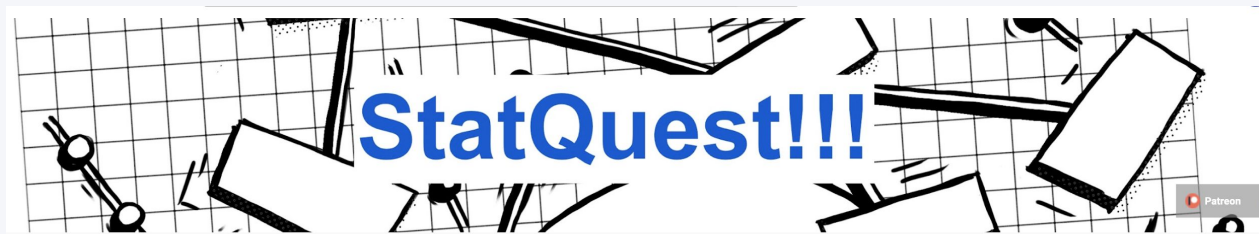
/

- Jogar uma moeda para o alto;
- Rolar um dado de 6 faces;
-

$$P(k|n,p) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$



A distribuição Binomial



StatQuest!!!

StatQuest with Josh Starmer ✓
245 mil inscritos

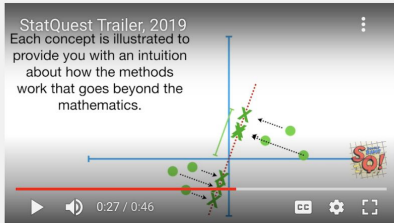
[SEJA MEMBRO](#) [INSCREVER-SE](#)

[INÍCIO](#) [VÍDEOS](#) [PLAYLISTS](#) [COMUNIDADE](#) [LOJA](#) [CANAIS](#) [SOBRE](#) [🔍](#)

StatQuest Trailer, 2019
Each concept is illustrated to provide you with an intuition about how the methods work that goes beyond the mathematics.

78.665 visualizações · 1 ano atrás

StatQuest breaks down complicated Statistics and Machine Learning methods into small, bite-sized pieces that are easy to understand. StatQuest doesn't dumb down the material, instead, it builds you up so that you are smarter and have a better understanding of Statistics and Machine Learning.



A series of overlapping geometric shapes, primarily triangles and quadrilaterals, in shades of teal and purple, located in the top-left corner of the slide.

AceLeraDev Data Science

Continuação Análise de dados exploratória

A series of overlapping geometric shapes, primarily triangles and quadrilaterals, in shades of teal and purple, located in the bottom-right corner of the slide.

Funções de PDF e CDF

/

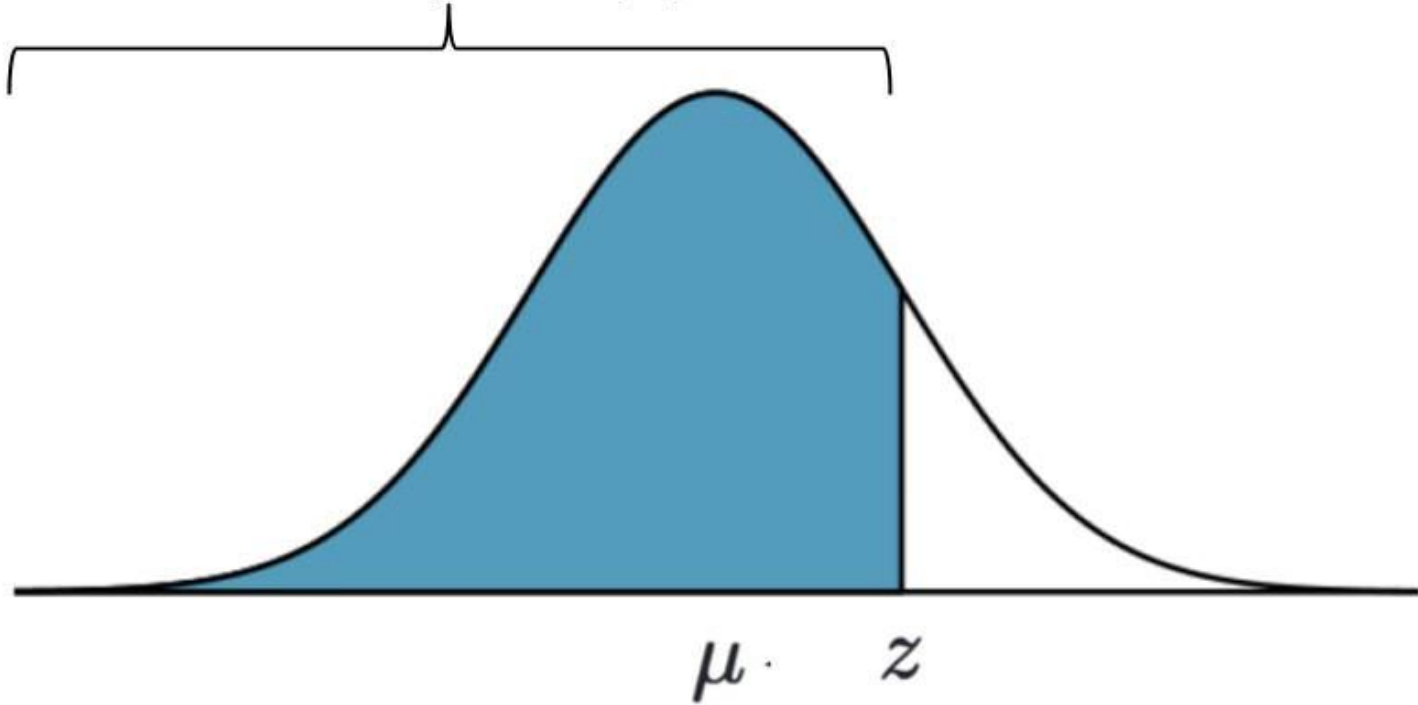
- Função Densidade de Probabilidade(PDF):
 - Descreve a probabilidade **relativa** de uma variável aleatória tomar um **valor dado**;
 - É não negativa sempre;
 - E sua integral sobre todo o espaço é igual a 1;
 - Informa a probabilidade de a variável X assumir um valor naquele intervalo.

Funções de PDF e CDF

/

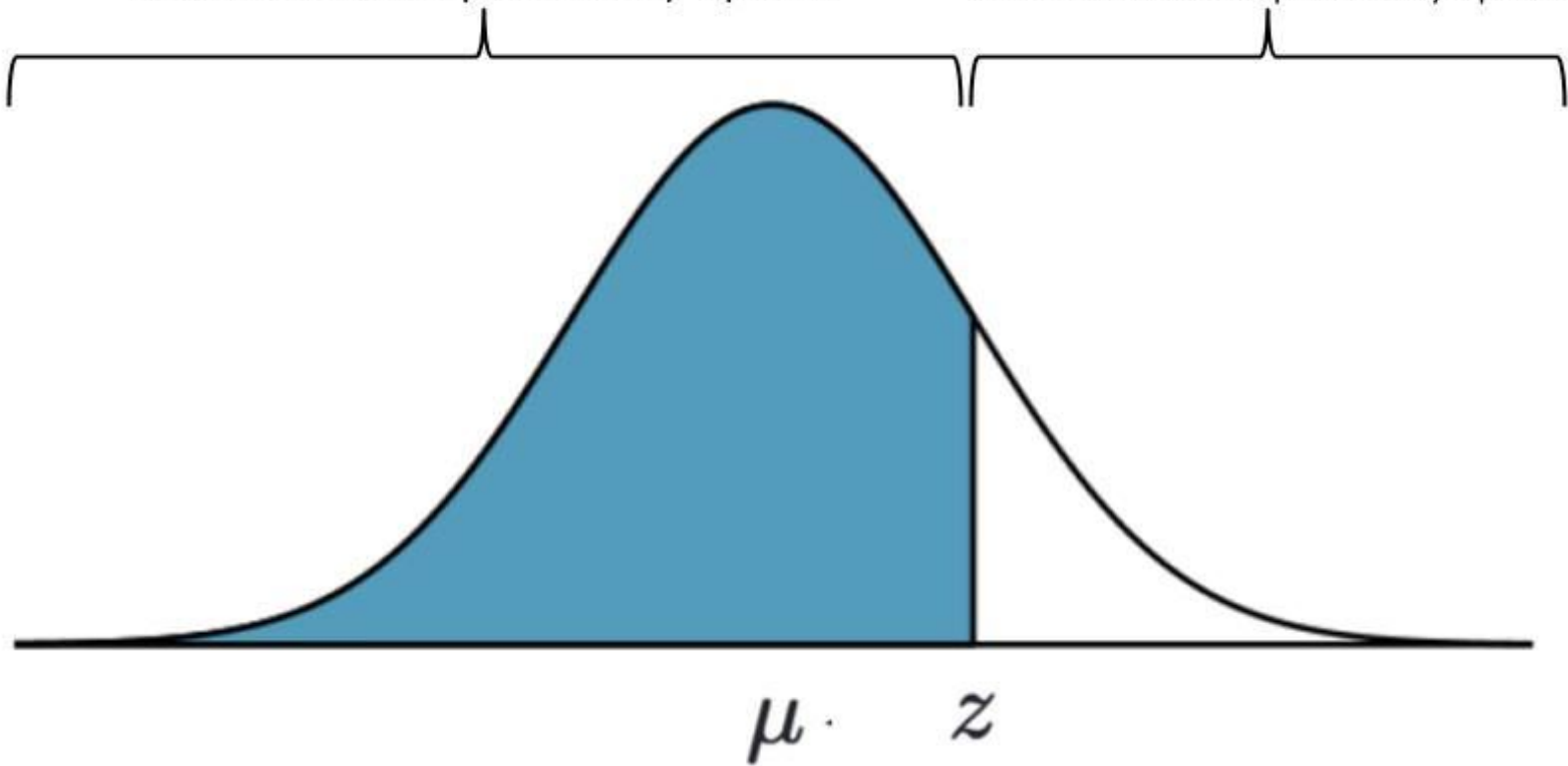
- Função Distribuição Acumulada(CDF):
 - Descreve a probabilidade **acumulada** de uma variável aleatória tomar um **conjunto de valores** dado;
 - É não negativa sempre;
 - E sua integral sobre todo o espaço é igual a 1;
 - Informa a probabilidade de a variável X assumir um valor naquele intervalo

The cumulative probability up to z

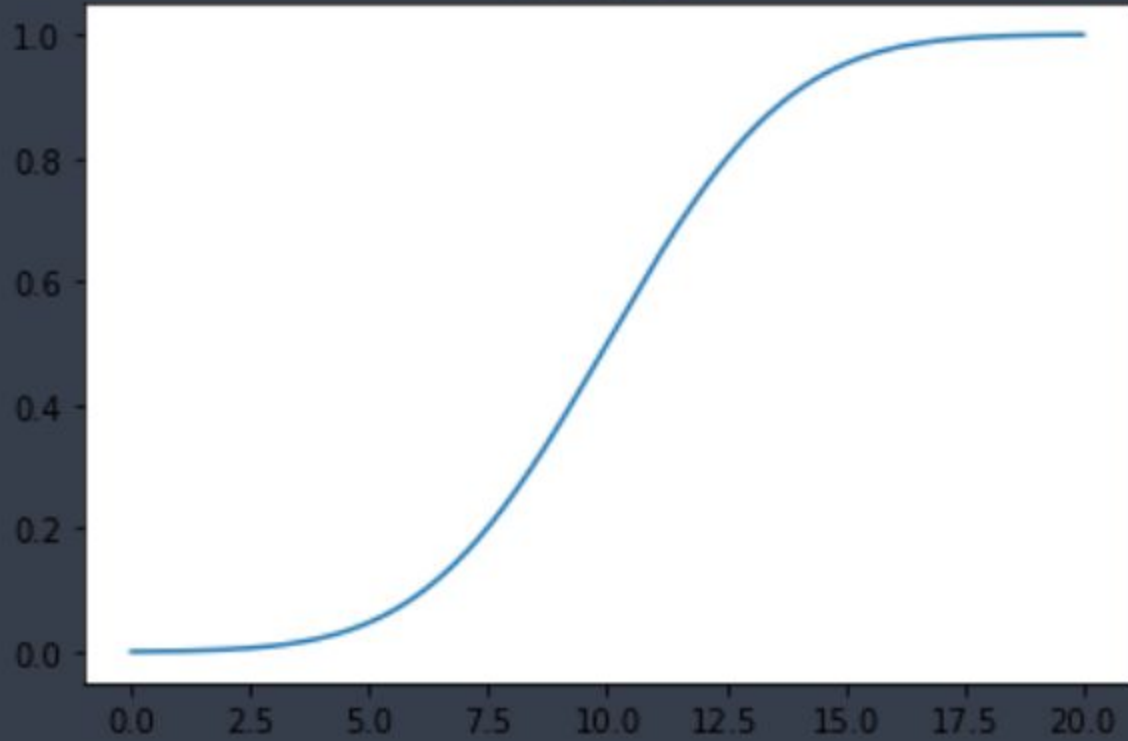


The cumulative probability up to z

1 - the cumulative probability up to z



Função CDF



Funções de PDF e CDF

/

- Função Distribuição Acumulada Empírica (ECDF):
 - Um estimador da função de distribuição cumulativa.;
 - Modelo **EMPÍRICO**;
 - **Funciona com suas observações**;

Links

/

- <https://brunomachadosilva.wordpress.com/2018/08/19/distribuicao-normal-python/>
- <https://pythonforundergradengineers.com/probability-under-normal-curve-with-python.html>
- <https://www.monolitonimbus.com.br/distribuicao-normal/>
- http://davidmlane.com/hyperstat/z_table.html
- <https://www.inf.ufsc.br/~andre.zibetti/probabilidade/normal.html>
- <http://www.portalaction.com.br/probabilidades/62-distribuicao-normal>
- https://pt.wikipedia.org/wiki/Distribui%C3%A7%C3%A3o_binomial
- <https://pt.khanacademy.org/math/statistics-probability/random-variables-stats-library/binomial-random-variables/v/visualizing-a-binomial-distribution>
- <http://ecologia.ib.usp.br/let/doku.php?id=tutoriais:pdf>
- <https://towardsdatascience.com/what-why-and-how-to-read-empirical-cdf-123e2b922480>