Entendendo Estatística Divertidamente

Profa. Adriana Silva

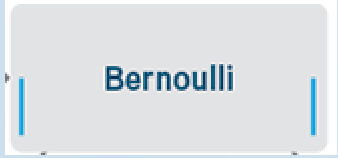
Seja bem vindX!!!
Câmera ligada e
Microfone mutado sempre
que não estiver falando

Distribuições de Probabilidade

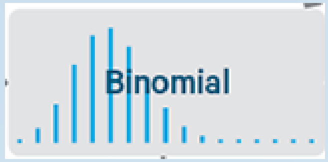
Variáveis Discretas

- Uniforme Discreta

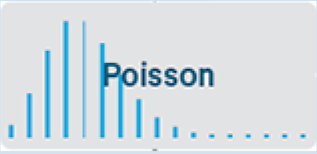
- Bernoulli



- Binomial



- Poisson

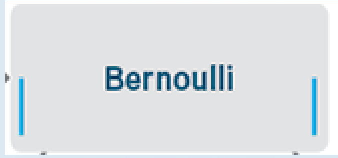


Distribuições de Probabilidade

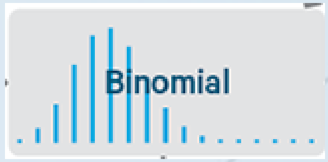
Variáveis Discretas

Uniforme Discreta

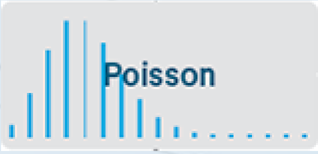
Uma das distribuições mais simples, é a distribuição uniforme, onde todos os valores possíveis possuem a mesma probabilidade de acontecerem.



Bernoulli



Binomial



Poisson

Data Science Analytics: uma visão geral

Distribuições de Probabilidade

Uniforme Discreta

- X: número de pontos obtidos no lançamento de um dado

X	1	2	3	4	5	6
P(X)	1/6	1/6	1/6	1/6	1/6	1/6

$$\mathrm{P}(X) = probabilidade\ de\ ocorrer\ o\ evento\ X$$

- Nesta distribuição todos os valores possuem a mesma probabilidade de ocorrer.
- Propriedade

$$\sum_{i=1}^{n} P(x_i) = 1 \;\; => \;\; \sum_{i=1}^{6} P(x_i) = 1$$

Distribuições de Probabilidade

Uniforme Discreta

X	P(X)	F(X)
1	0,167	0,167
2	0,167	0,334
3	0,167	0,500
4	0,167	0,667
5	0,167	0,834
6	0,167	1,000

$\mathrm{P}(X)$: *função de proababilidade*
$\mathrm{F}(X)$: *função de proababilidade acumulada*

$$\sum_{i=1}^{n} P(x_i) = 1$$

$$P(4) = \frac{1}{6}$$

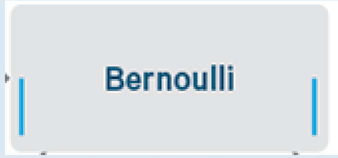
$$F(4) = P(X \leq 4) = P(1) + P(2) + P(3) + P(4) = 0{,}667$$

Distribuições de Probabilidade

Variáveis Discretas

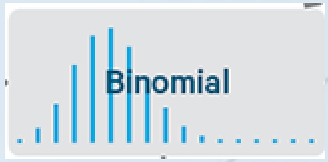
Uniforme Discreta

Uma das distribuições mais simples, é a distribuição uniforme, onde todos os valores possíveis possuem a mesma probabilidade de acontecerem.

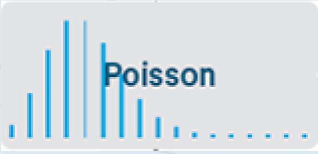


Bernoulli

Uma outra distribuição muito simples, usada para variáveis binárias (evento ocorrer ou não).



Binomial

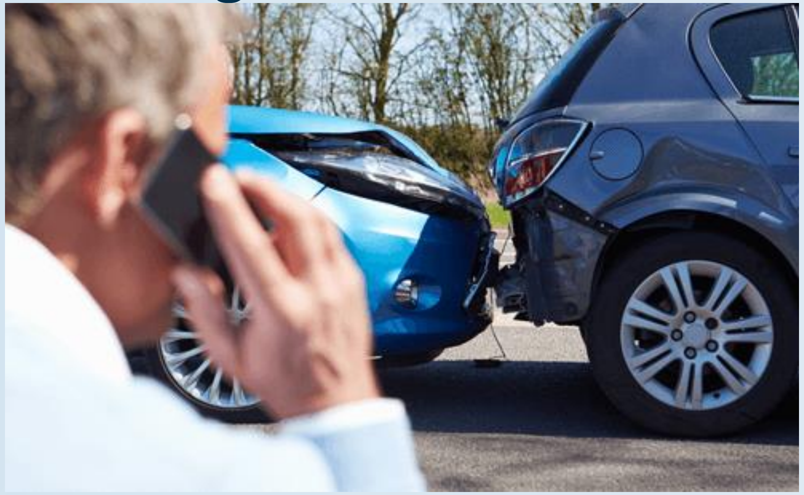


Poisson

Distribuições de Probabilidade

Bernoulli

- Um segurado é selecionado aleatoriamente do banco de dados de segurados de automóvel. O interesse é saber se o segurado sofreu algum tipo de sinistro.
- X=1 se o segurado sofreu sinistro
- X=0 se o segurado não sofreu sinistro
-
- P: probabilidade do segurado sofrer sinistro



X	0	1
P(X)	1-p	p

Distribuições de Probabilidade

Bernoulli

- Uma peça é selecionada aleatoriamente de um lote de peças. O interesse é saber se esta peça é defeituosa ou não.

  X=1 se a peça selecionada for defeituosa (sucesso)

  X=0 se a peça selecionada não for defeituosa (fracasso

P: probabilidade da peça selecionada ser defeituosa

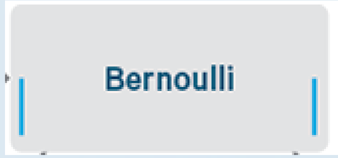
X	0	1
P(X)	1-p	p

Distribuições de Probabilidade

Variáveis Discretas

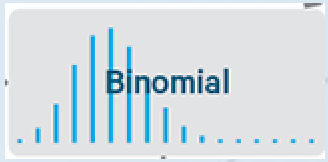
Uniforme Discreta

Uma das distribuições mais simples, é a distribuição uniforme, onde todos os valores possíveis possuem a mesma probabilidade de acontecerem.



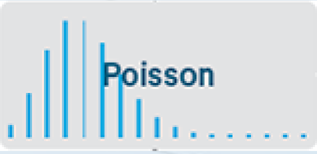
Bernoulli

Uma outra distribuição muito simples, usada para variáveis binárias (evento ocorrer ou não).



Binomial

Uma das distribuições mais mais utilizadas em toda a estatística. Com ela conseguimos calcular a probabilidade do número de vezes em que um evento ocorre.



Poisson

Distribuições de Probabilidade

Binomial

- Observa-se o comportamento diário de dois fundos de Investimentos (A e B). O retorno dos fundos são independentes.

- Os dois fundos possuem probabilidade de 0,10 de apresentar um comportamento de alta e 0,90 de apresentar um comportamento de baixa.

- Considerando um dia de operação qual a probabilidade de que os dois fundos apresentem alta?

Data Science Analytics: uma visão geral

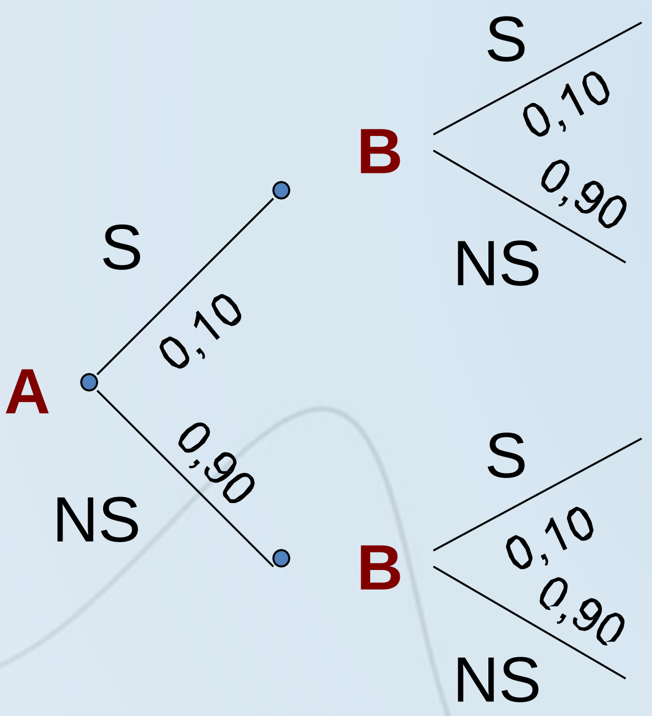
Distribuições de Probabilidade

Binomial

- Árvore de Probabilidade

S: Sucesso (alta do fundo)

X: Número de fundos em alta



Resultado	Valores de X	
S / S	2	(0,10)*(0,10) = 0,01
S / NS	1	(0,10)*(0,90) = 0,09
NS / S	1	(0,90)*(0,10) = 0,09
NS / NS	0	(0,90)*(0,90) = 0,81

Distribuições de Probabilidade

Binomial

- Árvore de Probabilidade

S: Sucesso (alta do fundo)

X: Número de fundos em alta

Resultado		Probabilidade	P(X)	X
S	S	(0,10)*(0,10)	0,01	2
S	NS	(0,10)*(0,90)	0,09	1
NS	S	(0,90)*(0,10)	0,09	1
NS	NS	(0,90)*(0,90)	0,81	0

X	P(X)
0	0,81
1	0,18
2	0,01

X : Número de fundos em alta

n : Tamanho da amostra – número de fundos

p : Probabilidade de alta do fundo

Distribuições de Probabilidade

Binomial

- Função de probabilidade

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

lembrando que

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

Distribuições de Probabilidade

Binomial

- Propriedades
  - O experimento consiste de uma seqüência de n ensaios idênticos;
  - Dois resultados são possíveis em cada ensaio;
  - A probabilidade de ocorrência do evento de interesse permanece constante em todos os ensaios;
  - Os ensaios são independentes;
  - Valores que a variável pode assumir: $X = 0,1,2,3,...,n$;

Distribuições de Probabilidade

Binomial – Exercício

- Considerando um dia de operação qual a probabilidade de que os dois fundos apresentem alta?

$$P(X = 2) = \binom{2}{2} 0{,}1^2 (0{,}9)^{2-2}$$

sendo que

$$\binom{2}{2} = \frac{2!}{2!\,(2-2)!} = \frac{2}{2} = 1$$

$$P(X = 2) = 0{,}1^2 (0{,}9)^0 = 0{,}01$$

Distribuições de Probabilidade

Binomial – Exercício

- Considerando um dia de operação qual a probabilidade de que um fundo apresente alta?

$$P(X = 1) = \binom{2}{1} 0{,}1^1 (0{,}9)^{2-1}$$

sendo que

$$\binom{2}{1} = \frac{2!}{1!\,(2-1)!} = \frac{2}{1} = 2$$

$$P(X = 1) = 2 * 0{,}1^1 * (0{,}9)^1 = 0{,}18$$

Data Science Analytics: uma visão geral

Distribuições de Probabilidade

Binomial

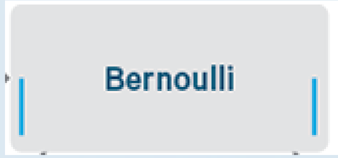
		p								
n	*x*	0,10	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50
2	0	0,8100	0,7225	0,6400	0,5625	0,4900	0,4225	0,3600	0,3025	0,2500
	1	0,1800	0,2550	0,3200	0,3750	0,4200	0,4550	0,4800	0,4950	0,5000
	2	0,0100	0,0225	0,0400	0,0625	0,0900	0,1225	0,1600	0,2025	0,2500
3	0	0,7290	0,6141	0,5120	0,4219	0,3430	0,2746	0,2160	0,1664	0,1250
	1	0,2430	0,3251	0,3840	0,4219	0,4410	0,4436	0,4320	0,4084	0,3750
	2	0,0270	0,0574	0,0960	0,1406	0,1890	0,2389	0,2880	0,3341	0,3750
	3	0,0010	0,0034	0,0080	0,0156	0,0270	0,0429	0,0640	0,0911	0,1250
4	0	0,6561	0,5220	0,4096	0,3164	0,2401	0,1785	0,1296	0,0915	0,0625
	1	0,2916	0,3685	0,4096	0,4219	0,4116	0,3845	0,3456	0,2995	0,2500
	2	0,0486	0,0975	0,1536	0,2109	0,2646	0,3105	0,3456	0,3675	0,3750
	3	0,0036	0,0115	0,0256	0,0469	0,0756	0,1115	0,1536	0,2005	0,2500
	4	0,0001	0,0005	0,0016	0,0039	0,0081	0,0150	0,0256	0,0410	0,0625
5	0	0,5905	0,4437	0,3277	0,2373	0,1681	0,1160	0,0778	0,0503	0,0312
	1	0,3280	0,3915	0,4096	0,3955	0,3602	0,3124	0,2592	0,2059	0,1562
	2	0,0729	0,1382	0,2048	0,2637	0,3087	0,3364	0,3456	0,3369	0,3125

Distribuições de Probabilidade

Variáveis Discretas

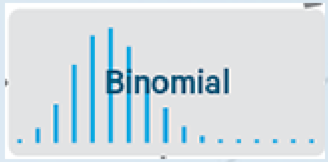
Uniforme Discreta

Uma das distribuições mais simples, é a distribuição uniforme, onde todos os valores possíveis possuem a mesma probabilidade de acontecerem.



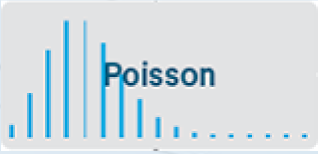
Bernoulli

Uma outra distribuição muito simples, usada para variáveis binárias (evento ocorrer ou não).



Binomial

Uma das distribuições mais mais utilizadas em toda a estatística. Com ela conseguimos calcular a probabilidade do número de vezes em que um evento ocorre.



Poisson

Frequentemente utilizada para estimar o número de ocorrências ao longo de um intervalo de tempo oou espaço.

Distribuições de Probabilidade

Poisson

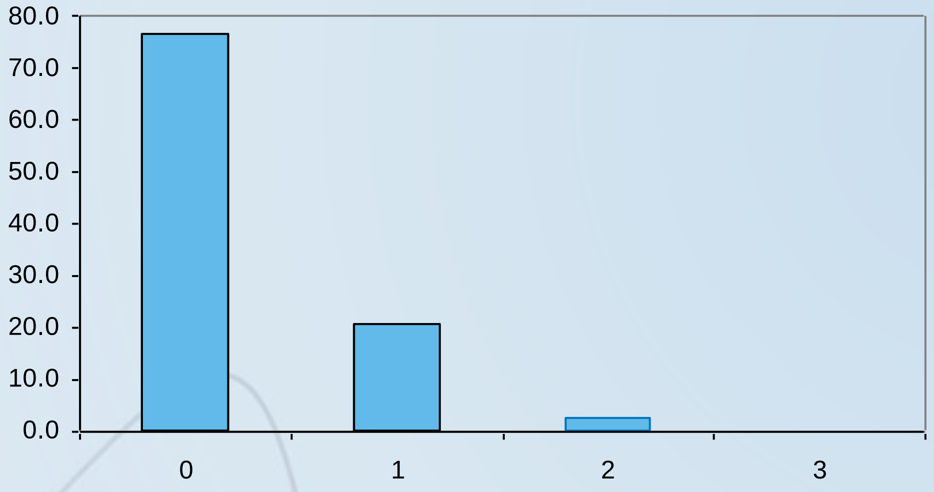
- A distribuição de Poisson é empregada quando desejamos contar o número de eventos de certo tipo que ocorrem num intervalo pré-determinado (tempo, medida linear, área, volume, etc).

  - Número de chamadas telefônicas recebidas em um intervalo de cinco minutos;
  - Número de falhas de um sistema bancário em um dia de operação;
  - Número de acidentes ocorridos em um dia na cidade de São Paulo;
  - Número de defeitos para cada 100 metros de tecido;

Distribuições de Probabilidade

Poisson

- Variável em estudo: número de sinistros no período de 12 meses



Média 0,25 sinistros por ano

$$Poisson\ (0{,}25)$$

Distribuições de Probabilidade

Poisson

- Função de Probabilidade de uma variável aleatória com distribuição de Poisson

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

sendo e = 2,71828

$$\lambda = número\ médio\ da\ variável\ aleatória\ no\ período\ de\ estudo$$

Distribuições de Probabilidade

Poisson – Exercício

- Qual a probabilidade de obtermos 8 sinistros durante o período de um ano sabendo-se o número médio de sinistrados no período de um ano é 6?

$$P(X = k) = \frac{e^{-\lambda}\lambda^{k}}{k!}$$

$$P(X = 8) = \frac{e^{-6}6^{8}}{8!} = 0{,}1032$$

Distribuições de Probabilidade

Poisson – Exercício

- Qual a probabilidade de obtermos no máximo 8 sinistros durante o período de um ano sabendo-se o número médio de sinistrados no período de um ano é 6 ?

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

$$P(X \leq 8) = \text{P(X=0)+ P(X=1)+ P(X=2)+ P(X=3)+}$$
$$\text{P(X=4)+ P(X=5)+ P(X=6)+ P(X=7)+ P(X=8)}$$

Distribuições de Probabilidade

Poisson – Exercício

- Qual a probabilidade de obtermos 2 sinistros durante o período de 4 meses sabendo-se que o número médio de sinistrados no período de um ano é 6?

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

Se são 6 sinistrados em um ano, é o mesmo que 2 em 4 meses

$$P(X = 2) = \frac{e^{-2}2^2}{2!} = 0{,}2706$$

Probabilidade

Teorema de Monty Hall

Distribuições de Probabilidade

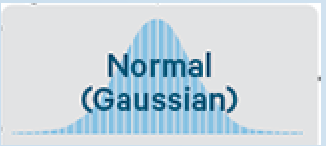
Variáveis Contínuas

- Em uma variável contínua, a probabilidade da variável aleatória assumir um valor dentro de um dado intervalo [x1 ; x2] é definido pela área abaixo da curva da função densidade de probabilidades entre x1 e x2.

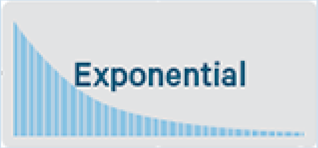
Distribuições de Probabilidade

Variáveis Contínuas

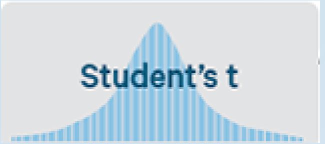
- Normal



- Exponencial



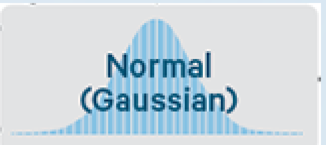
- t de Student



Data Science Analytics: uma visão geral

Distribuições de Probabilidade

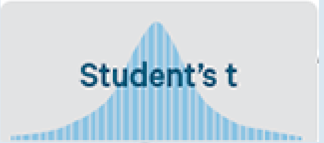
Variáveis Contínuas



Normal

A distribuição normal é uma das distribuições de probabilidade mais utilizadas para modelar fenômenos naturais. Isso se deve ao fato de que um grande número de fenômenos naturais apresenta sua distribuição de probabilidade tão proximamente normal. A distribuição normal também é chamada distribuição gaussiana.

Exponencial

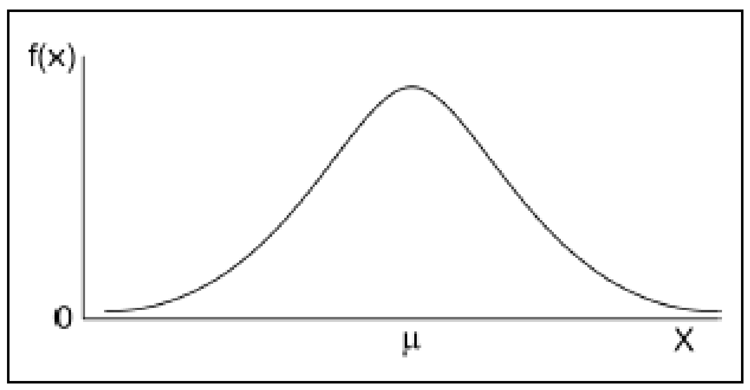


t de Student

Distribuições de Probabilidade

Normal

- A **Distribuição Normal,** também conhecida como Distribuição de Gauss ou Gaussiana foi primeiramente introduzida pelo Matemático Abraham de Moivre em 1733. Foi apresentada no contexto da aproximação de **Distribuições Binomiais** para grandes valores amostrais.
- O formato, ou forma, da **Distribuição Normal de Probabilidade** é ilustrado pela curva em forma de sino.



$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\frac{-(x-\mu)^2}{2\sigma^2}$$

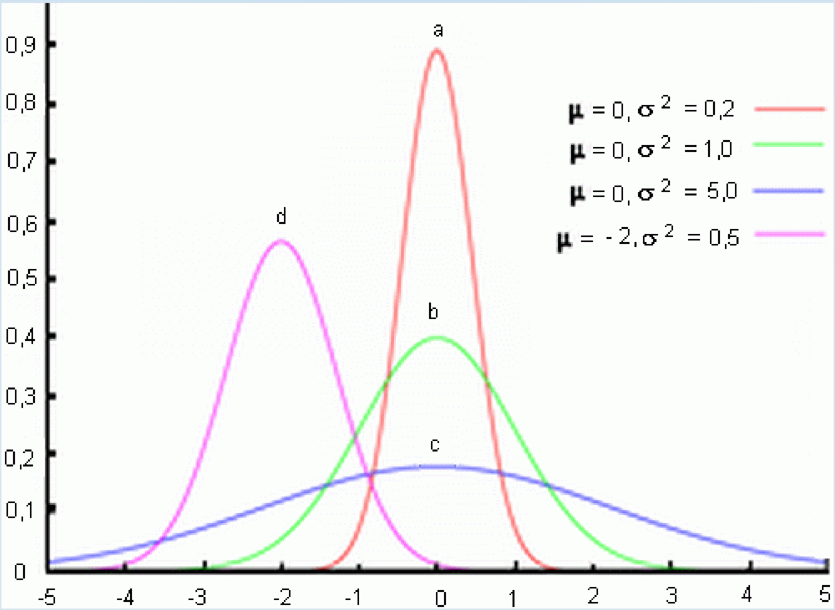
onde $-\infty < x < \infty$ e $\sigma > 0$

Distribuições de Probabilidade

Normal

- A família das Distribuições Normais de Probabilidade é diferenciada por dois parâmetros: sua **Média** e seu **Desvio Padrão (Variância)**.

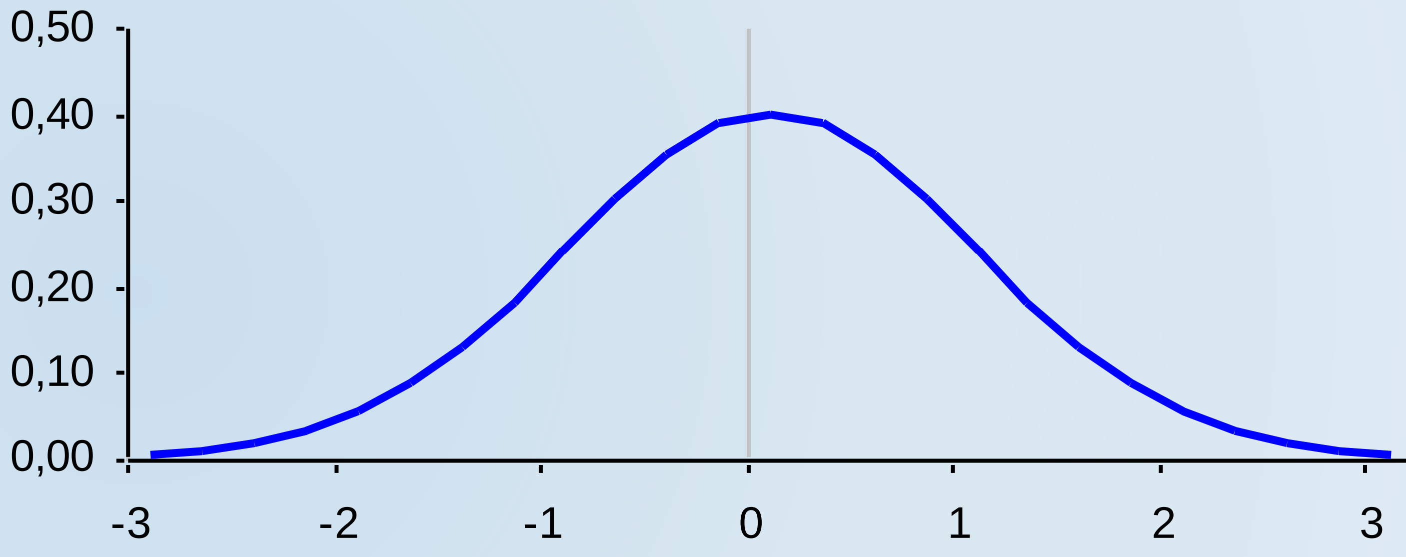
- A **Média** pode ser qualquer valor numérico: negativo, zero ou positivo e o **Desvio Padrão** determina quanto uma curva é achatada ou larga.



Distribuições de Probabilidade

Normal

- Algumas características importantes:
  - Simétrica em torno da média
  - A área sob a curva é 1
  - Média = Mediana = Moda



Distribuições de Probabilidade

Normal

- A Função de Densidade Normal é dita como Função de Densidade Normal **Padrão** quando a média é igual a 0 e o Desvio Padrão é igual a 1 e a fórmula é dada por:

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

$$\mathrm{f}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

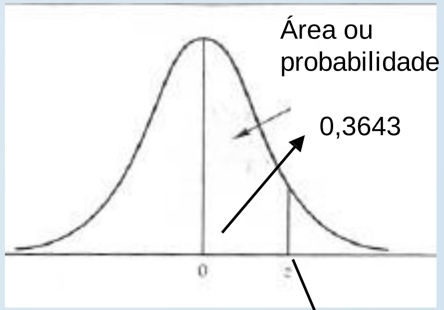
onde $-\infty < x < \infty$ e $\sigma > 0$

- A maioria dos livros de estatística possuem tabelas com os valores de probabilidades para cada valor possível de x para a distribuição normal padrão N(0,1).

Data Science Analytics: uma visão geral

Distribuições de Probabilidade

Normal



1,1

P(0 < Z < 1,1) = 0,3643

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2518	0,2549
0,7	0,2580	0,2612	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2.3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2.4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2.5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2.6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2.7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2.8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2.9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3.0	0,4986	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990

Distribuições de Probabilidade

Normal – Exercício

- Supondo que Z segue uma distribuição normal padrão [$Z \sim N(0,1)$]. Desenhe a curva e calcule:

1) $P(0 < Z < 1{,}96)$
2) $P(-1{,}64 < Z < 0)$
3) $P(Z > 1{,}62)$
4) $P(Z < -0{,}84)$
5) $P(-0{,}68 < Z < 1{,}85)$

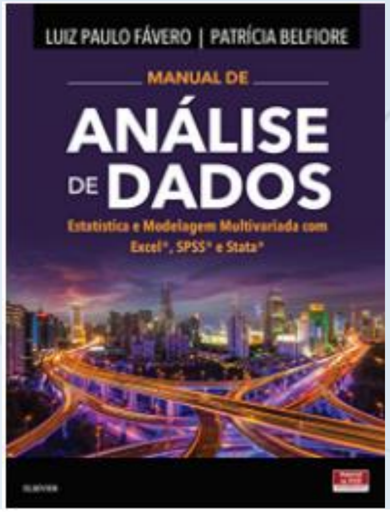
6. $P(-1 < Z < 1)$
7. $P(-2 < Z < 2)$
8. $P(-3 < Z < 3)$

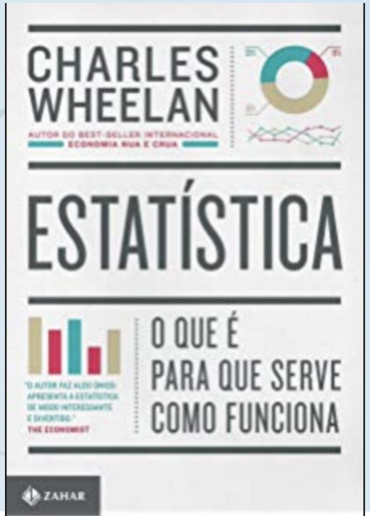
Não esqueça de deixar seu feedback!

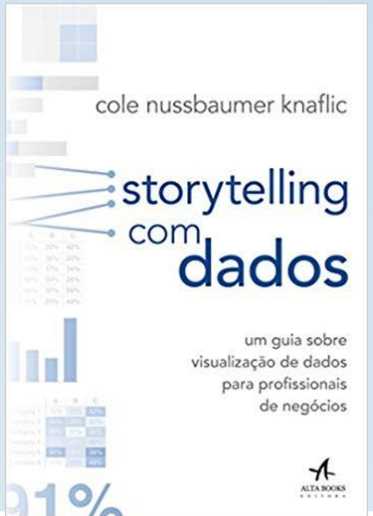
=]

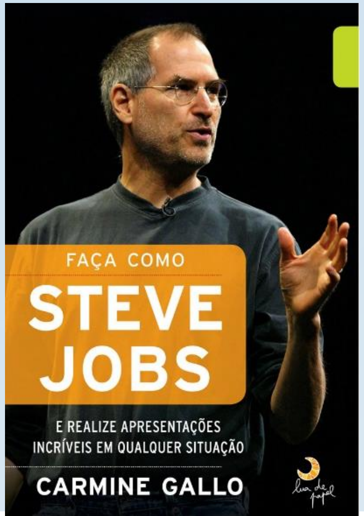
Referência

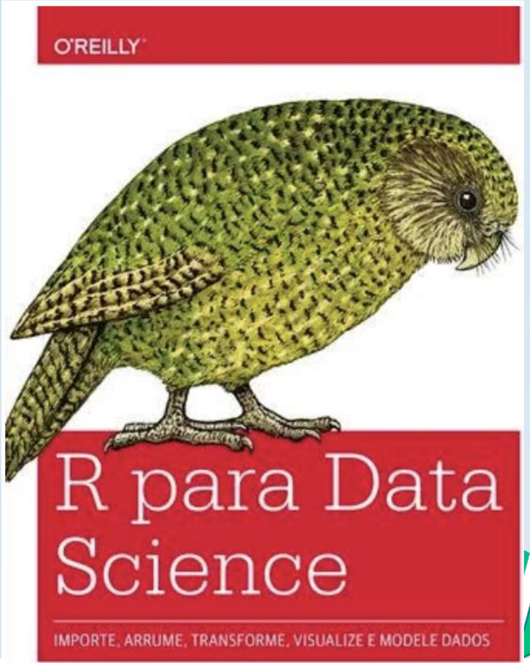
- Moore, D., McCabe, G., Duckworth, W., Sclove, S. *A prática da Estatística Empresarial*. LTC, Rio de Janeiro, 2006.
- Anderson, D., Sweeney, D., Williams, T. *Estatística Aplicada à Administração e Economia*. Segunda Edição. Cengage Learning, São Paulo, 2011.
- www.asn.rocks
- www.curso-r.com











Data Science Analytics: uma visão geral

It´s kind of fun to do the IMPOSSIBLE

dri@asn.rocks

/in/adrianamms
/in/asn.rocks

asn.rocks

www.asn.rocks