Analytics & Inteligência Artificial

Tema da aula
Distribuições de Probabilidades & Inferência Estatística

LABDATA | FIA FUNDAÇÃO INSTITUTO DE ADMINISTRAÇÃO BUSINESS SCHOOL

FIA Business School

NOSSOS DIFERENCIAIS | QUEM SOMOS

BUSINESS SCHOOL
Graduação, pós-graduação, MBA, Pós- MBA, Mestrado Profissional, Curso In *Company* e EAD

CONSULTING
Consultoria personalizada que oferece soluções baseada em seu problema de negócio

RESEARCH
Atualização dos conhecimentos e do material didático oferecidos nas atividades de ensino

FUNDAÇÃO INSTITUTO DE ADMINISTRAÇÃO

Líder em Educação Executiva, referência de ensino nos cursos de graduação, pós-graduação e MBA, tendo excelência nos programas de educação. Uma das principais escolas de negócio do mundo, possuindo convênios internacionais com Universidades nos EUA, Europa e Ásia. +8.000 projetos de consultorias em organizações públicas e privadas.

NOTA MÁXIMA ENADE 2019 MELHOR ESCOLA DO BRASIL

Único curso de graduação em administração a receber as notas máximas

A primeira escola brasileira a ser finalista da maior competição de MBA do mundo

Única *Business School* brasileira a figurar no *ranking* LATAM

Signatária do Pacto Global da ONU

Membro fundador da ANAMBA - Associação Nacional MBAs

Credenciada pela AMBA - *Association of MBAs*

Credenciada ao *Executive MBA Council*

Filiada a AACSB - *Association to Advance Collegiate Schools of Business*

Filiada a EFMD - *European Foundation for Management Development*

Referência em cursos de MBA nas principais mídias de circulação

O **Laboratório de Análise de Dados** – LABDATA é um Centro de Excelência que atua nas áreas de ensino, pesquisa e consultoria em análise de informação utilizando técnicas de ***Big Data***, ***Analytics*** e **Inteligência Artificial**.

Profª Drª Alessandra

O LABDATA é um dos pioneiros no lançamento dos cursos de *Big Data* e *Analytics* no Brasil

Os diretores foram professores de grandes especialistas do mercado

+10 anos de atuação

+1000 alunos formados

Docentes

- Sólida formação acadêmica: doutores e mestres em sua maioria
- Larga experiência de mercado na resolução de *cases*
- Participação em Congressos Nacionais e Internacionais
- Professor assistente que acompanha o aluno durante todo o curso

Estrutura

- 100% das aulas realizadas em laboratórios
- Computadores para uso individual durante as aulas
- 5 laboratórios de alta qualidade (investimento +R$2MM)
- 2 Unidades próximas a estação de metrô (com estacionamento)

CONTEÚDO PROGRAMÁTICO

ANALYTICS

MACHINE LEARNING

DEEP LEARNING

ANÁLISE EXPLORATÓRIA

TÉCNICAS DE PROJEÇÃO

TÉCNICAS DE CLASSIFICAÇÃO

TÉCNICAS DE SEGMENTAÇÃO

TÉCNICAS DE ANALYTICS

LINGUAGEM

PYTHON

R

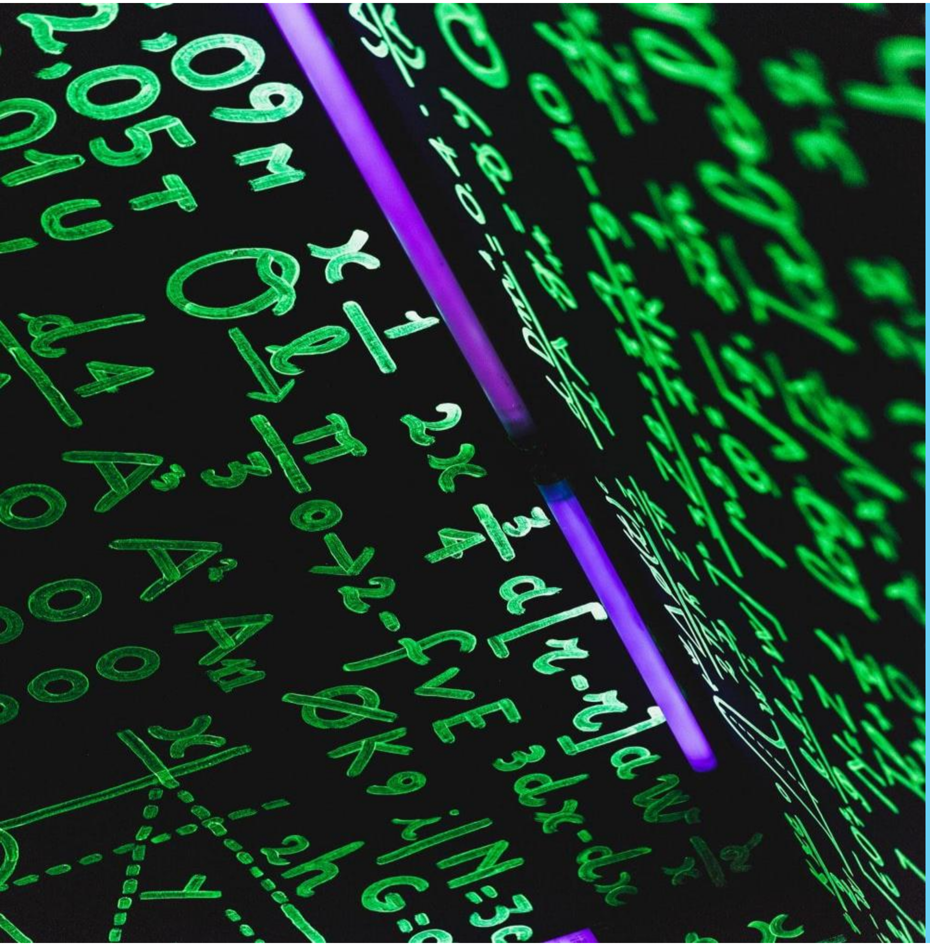
PROJETO

PROJETO ANALYTICS

Conteúdo da Aula

1. Distribuições de Probabilidades
   i. Distribuições Discretas
      a. Uniforme
      b. Bernoulli
      c. Binomial
   ii. Distribuições Contínuas
      a. Normal
      b. T-*student*
2. Inferência Estatística
      a. Conceitos Preliminares
      b. Distribuição Amostral da Média – Teorema do Limite Central
      c. Aplicações
3. Exercícios

1. Distribuições de probabilidades



Na Análise Exploratória de Dados, aprendemos como analisar um determinado conjunto de dados por meio de técnicas gráficas e medidas resumo para entender a distribuição dos dados.

Com suposições adequadas e sem observarmos diretamente o fenômeno de interesse, podemos criar um modelo teórico que reproduza de maneira razoável a distribuição dos dados.

Tais modelos são chamados de modelos probabilísticos ou **distribuições de probabilidades** que serão objetos de estudo desta aula.

Variável aleatória

Definição: uma variável aleatória (v.a.) é uma variável cujo valor é o resultado numérico de um experimento.

Experimento: inspecionar a fabricação de 50 televisores.

X : número de televisores com defeito.

Possíveis valores de X={0,1, 2, 3, …., 50}.

Note que uma variável é uma característica da unidade amostral, por exemplo, tamanho da TV, peso da TV, modelo da TV, etc.

Variável aleatória discreta

1.i DISTRIBUIÇÕES DISCRETAS | DISTRIBUIÇÕES DE PROBABILIDADES

Uma variável aleatória discreta pode assumir um número finito ou infinito de **valores contáveis**.

- Variável aleatória discreta com **número finito** de valores:

  x = número de carros de luxo vendidos por dia,
  x assume somente 5 valores (0, 1, 2, 3, 4).

- Variável aleatória discreta com **número infinito** de valores:

  x = número de carros que passam por uma determinada rua,
  x assume valores 0, 1, 2, 3, ...

  Observação: Podemos contar os carros que chegam, mas não há limite superior finito sobre o número que pode chegar.

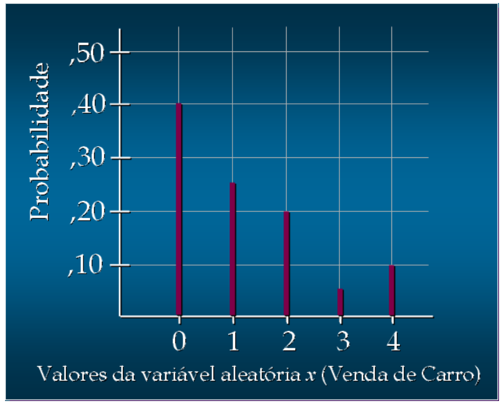
A função de probabilidade é denotada por *f (x)*, que fornece a probabilidade de cada valor da variável aleatória.

Usando o exemplo dos carros de luxo:

Unidades Vendidas	Número de Dias
0	80
1	50
2	40
3	10
4	20
	200

80/200

x	*f(x)*
0	0,40
1	0,25
2	0,20
3	0,05
4	0,10
	1,00



A função de probabilidade acumulada é denotada por F(x), que fornece a probabilidade acumulada de cada valor da variável aleatória.

Continuando o exemplo dos carros de luxo:

Unidades Vendidas	Número de Dias
0	80
1	50
2	40
3	10
4	20
	200

f(X): Função de probabilidade RELATIVA

x	f(x)
0	0,40
1	0,25
2	0,20
3	0,05
4	0,10
	1,00

F(X): Função de probabilidade ACUMULADA

x	F(x)
0	0,40
1	0,65
2	0,85
3	0,90
4	1,00

Variável aleatória discreta

A literatura estatística apresenta várias **distribuições de probabilidades discretas**.

A seguir, apresentaremos as distribuições mais utilizadas no mercado e que serão utilizadas em técnicas e modelos estatísticos mais avançados ao longo deste curso.

Uniforme

Bernoulli

Binomial

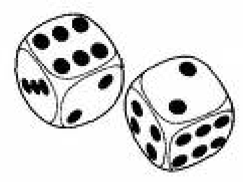
Distribuição Uniforme

Todos os valores possuem a **mesma probabilidade** de ocorrer.

Exemplo:

X: número da face no lançamento de um dado.

x	1	2	3	4	5	6
$f(x)$	1/6	1/6	1/6	1/6	1/6	1/6



probabilidade de ocorrer x

Todos os valores possuem a **mesma probabilidade** de ocorrer.

Aplicação *Business:*

Em uma operação de telemarketing o gerente de vendas, de acordo com último relatório de vendas identificou que os 4 tipos de seguros de emergências médicas, nos valores mensais de prêmio R$9,90, R$14,90, 19,90 e R$29,90, a serem oferecidos na operação possuem a mesma probabilidade de venda.

X: valor do seguro da emergência médica.

x	9,9	14,9	19,9	29,9
$f(x)$	0,25	0,25	0,25	0,25

probabilidade de ocorrer a venda de x

Distribuição Bernoulli

O resultado da variável aleatória assume somente 2 valores: ocorrência ou não de uma determinada característica.

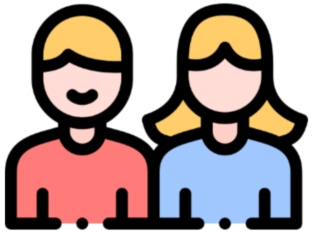
Em que, **p** é a **probabilidade do evento de interesse** ocorrer, e **1-p** é a **probabilidade do evento complementar**.

Exemplos:

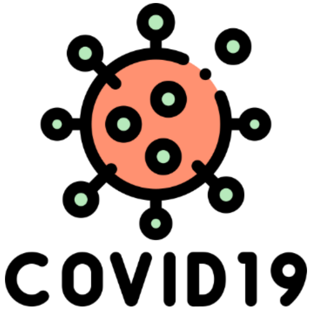
Lançamento de uma moeda.

Avaliar se uma peça é defeituosa.

Sexo do indivíduo escolhido para uma pesquisa.



Inquérito de saúde: testou positivo para a Covid-19.



Distribuição Bernoulli

O resultado da variável aleatória assume somente 2 valores: ocorrência ou não de uma determinada característica.

Em que, **p** é a **probabilidade do evento de interesse** ocorrer, e **1-p** é a **probabilidade do evento complementar**.

Aplicação *Business*:

Uma empresa de telefonia móvel vem sofrendo problemas técnicos e muitos clientes estão solicitando cancelamento do serviço.

Temos o interesse em saber se o cliente ligou para ilha de retenção para solicitar cancelamento da linha.

X=1 se o cliente ligou para ilha de retenção,

X=0 se o cliente não ligou para ilha de retenção.

Distribuição Bernoulli

O resultado da variável aleatória assume somente 2 valores: ocorrência ou não de uma determinada característica.

Em que, **p** é a **<u>probabilidade do evento de interesse</u>** ocorrer, e **1-p** é a **probabilidade do evento complementar**.

Aplicação *Business*:

Um banco tem o interesse em analisar clientes com alto valor de fundo de investimento.

Temos o interesse em saber se o cliente atingiu mais de R$100.000 investido.

X=1 se o cliente tem >=R$100.000 investido,

X=0 se o cliente <R$100.000 investido.

1. Uma loja de informática faz diariamente vendas de microcomputadores e acessórios. Foi avaliado no relatório do mês anterior que 70% das vendas foram de acessórios.

(a) Qual a variável aleatória?
(b) Qual a distribuição de probabilidades é adequada para modelar este problema?
(c) O gerente deseja saber se a próxima venda será um micro ou um acessório. Com a maior probabilidade, essa venda será um microcomputadores ou acessórios?

Um total de n indivíduos são observados e o resultado do evento para cada indivíduo assume somente 2 valores: ocorrência ou não de uma determinada característica.

p é a **<u>probabilidade do evento de interesse</u>** ocorrer, e **1-p** é a **probabilidade do evento complementar**.

A variável aleatória, neste caso, é a quantidade de indivíduos que assumem esta determinada característica.

Aplicação *Business*:

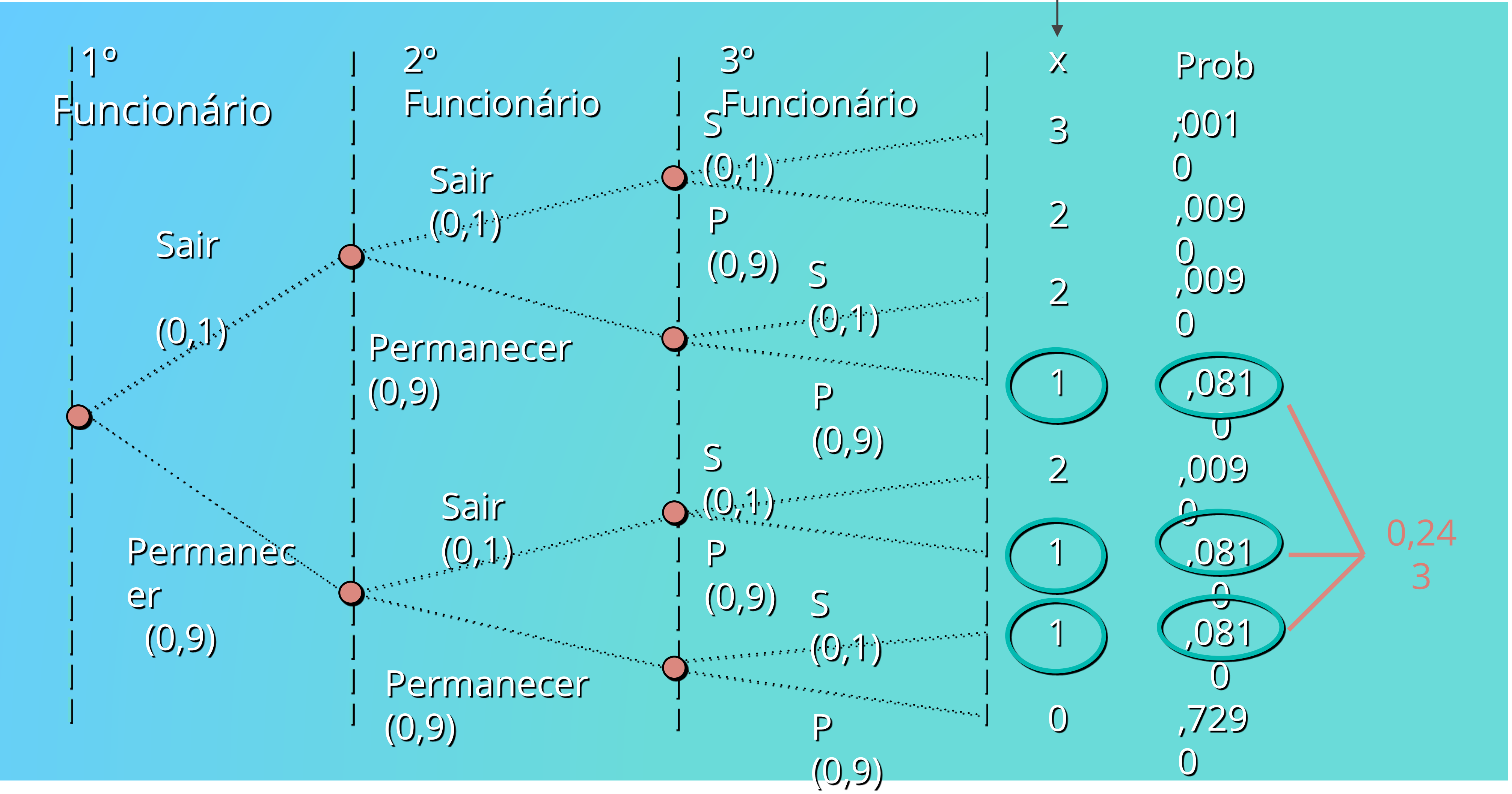
Uma empresa está preocupada com a taxa de retenção de seus funcionários. No último ano, o RH da empresa observou uma taxa de demissão voluntária de 10%.

Assim, para qualquer funcionário escolhido aleatoriamente, o RH estima uma probabilidade de 0,1 que a pessoa não estará na empresa no próximo ano.

Escolhendo 3 funcionários da empresa ao acaso, qual é a probabilidade de 1 deles deixar a companhia no próximo ano?

Distribuição Binomial

X: quantidade de funcionários que saem da empresa



Usando a tabela da distribuição Binomial:

		p									
n	**x**	.05	,10	,15	,20	,25	,30	,35	,40	,45	,50
3	0	,8574	,7290	,6141	,5120	,4219	,3430	,2746	,2160	,1664	,1250
	1	,1354	,2430	,3251	,3840	,4219	,4410	,4436	,4320	,4084	,3750
	2	,0071	,0270	,0574	,0960	,1406	,1890	,2389	,2880	,3341	,3750
	3	,0001	,0010	,0034	,0080	,0156	,0270	,0429	,0640	,0911	,1250

Escolhendo 3 funcionários da empresa ao acaso, a probabilidade de 1 deles deixar a companhia no próximo ano é de 24,3%.

Usando Excel:

	A	B
1	*x*	*f(x)*
2	1	0,24300

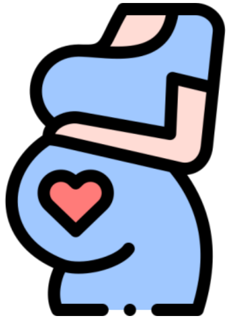
=DISTRBINOM(A2;3;0,1;FALSO)

DISTRBINOM(núm_s;tentativas;probabilidade_s;cumulativo)

Um total de **n indivíduos** são observados e o resultado da variável aleatória para cada indivíduo assume somente 2 valores: **ocorrência <u>ou</u> não** de uma determinada característica.

Em que, **p** é a **<u>probabilidade do evento de interesse</u>** ocorrer, e **1-p** é a **probabilidade do evento complementar**.

Aplicação *Business*:



Um administrador de um hospital baseado em sua experiência, sabe que 5% de todas os grávidas admitidas num pronto socorro (PS) devem ser colocadas para ter acompanhamento da equipe médica especializada em complicações de parto.

Use este dado para estimar a probabilidade de que em 120 gestantes novas admitidas em um dia no PS, 3 precisem da equipe média especializada.

A variável aleatória, neste caso, é a quantidade de novas gestantes admitidas em um dia no PS.

n=120

p=0,05

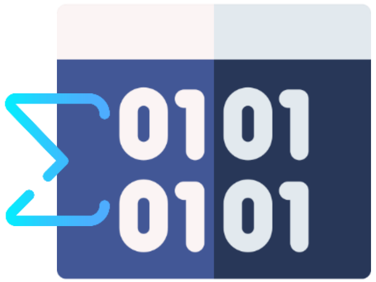
Probabilidade do evento de interesse.

	A	B	C	D	E	F	G
1							
2		**Binomial**					
3		*x*		*f(x)*			
4		3		0,087	=DISTRBINOM(B4;120;0,05;FALSO)		
5							

A probabilidade de que 3 pacientes novas precisem da equipe médica especializada dentre as 120 admitidas é 8,7%

Propriedades da Binomial

1. O experimento consiste em uma sequência de n tentativas idênticas.
2. Dois resultados, sucesso e falha, são possíveis em cada tentativa, ou seja, cada tentativa é uma Bernoulli.
3. A probabilidade de sucesso, denotada por p, não muda de tentativa para tentativa.
4. As tentativas são independentes entre si.



1. Em uma pesquisa sobre o uso de cartão de crédito, foi identificado que 9% dos estudantes universitários possuem limite de cartão de crédito acima de R$5.000. Suponha que 10 estudantes universitários sejam escolhidos aleatoriamente para uma entrevista em relação ao uso do cartão de crédito.

(a) Qual a variável aleatória?
(b) Qual a distribuição de probabilidade é adequada para modelar este problema?
(c) Qual a probabilidade de nenhum estudante ter limite de crédito superior a R$5.000?
(d) Qual a probabilidade de pelo menos três estudantes terem limites de crédito maiores do que R$5.000?

2. Considere um vendedor de seguros que visita 2 famílias selecionadas aleatoriamente. O resultado associado com cada visita está classificado como uma compra ou não de uma apólice de seguro residencial. Da experiência passada, o vendedor sabe que a probabilidade de que uma família selecionada aleatoriamente compre uma apólice de seguros é de 0,40 (quatro casas decimais).

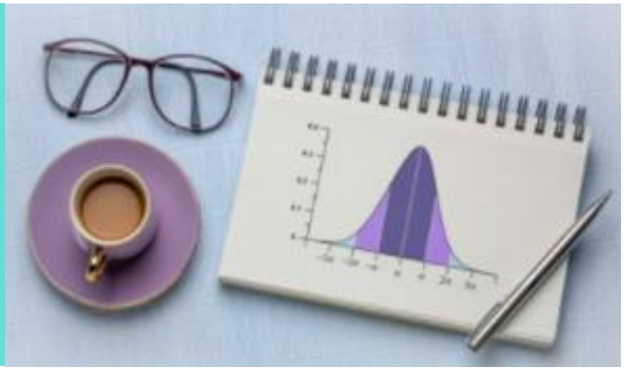
a) Qual o valor de n e qual o valor de p?
b) Obtenha a árvore Binomial de possibilidades.
c) Complete a tabela abaixo:

X	p
P(X = 0)	
P(X = 1)	
P(X = 2)	

A literatura estatística apresenta várias **distribuições de probabilidades contínuas**.

A seguir, apresentaremos as distribuições mais utilizadas no mercado e que serão utilizadas em técnicas e modelos estatísticos mais avançados ao longo deste curso.

- **Normal**
- **T-student**



Uma variável aleatória contínua pode assumir valores no **intervalo** dos números **reais**.

Exemplos:

Nota de uma prova

Volume de óleo trocado para motor de carro

Quantidade de chuva mm³

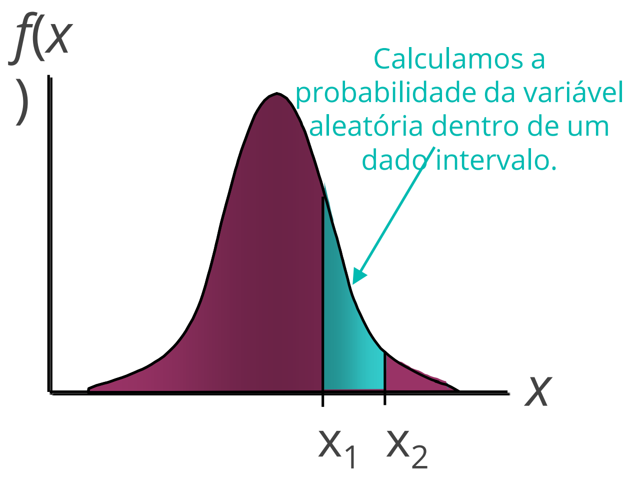
Uma variável aleatória contínua pode assumir valores no **intervalo** dos números **reais**.

Notação:

- f(x): função densidade (curva)
- F(x) = P(X<x): função de distribuição (área abaixo da curva) e assume valores entre [0;1].

A área total abaixo da curva sempre assumirá valor igual a 1.

A probabilidade da variável aleatória assumindo um valor dentro de um dado intervalo [x_1; x_2] é definido pela área abaixo da curva da função densidade de probabilidades entre x_1 e x_2.



Uma *startup* deseja definir um piso salarial para a carreira de cientista de dados dentro de sua empresa. Assim, decidiu realizar uma pesquisa salarial para estudar o salário anual destes profissionais no primeiro ano de sua carreira, na cidade de São Paulo, e em empresas com faturamento entre 5 e 20 milhões de reais ao ano.

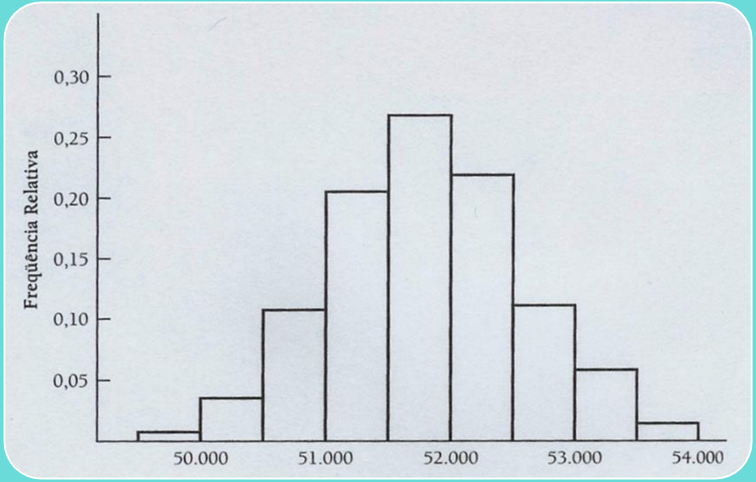
Estratégia da Pesquisa:

Contratou uma consultoria especializada em pesquisa de mercado que entrevistou o RH de 500 empresas, na cidade de São Paulo, com faturamento entre 5 e 20 milhões de reais por ano, e coletou a informação do salário anual dos cientistas de dados no primeiro ano carreira.

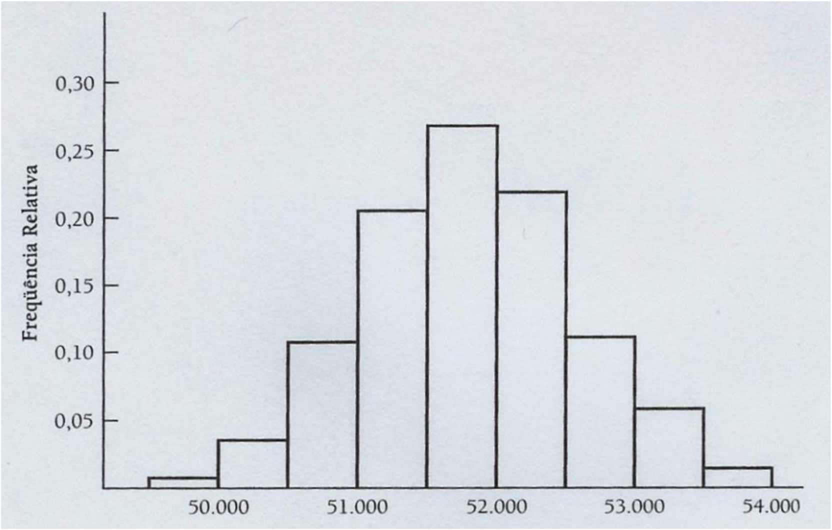
Exemplo: Pesquisa salarial

Empresa	Salário Anual
1	R$ 51.814,00
2	R$ 52.669,70
3	R$ 51.780,30
4	R$ 51.587,90
.	.
.	.
.	.
500	R$ 51.752,00

Salário anual	Freqüência Absoluta	Freqüência Relativa
49.500,00 a 49.999,99	2	0,004
50.000,00 a 50.499,99	16	0,032
50.500,00 a 50.999,99	52	0,104
51.000,00 a 51.499,99	101	0,202
51.500,00 a 51.999,99	133	0,266
52.000,00 a 52.499,99	110	0,220
52.500,00 a 52.999,99	54	0,108
53.000,00 a 53.499,99	26	0,052
53.500,00 a 53.999,99	6	0,012
Total	500	1

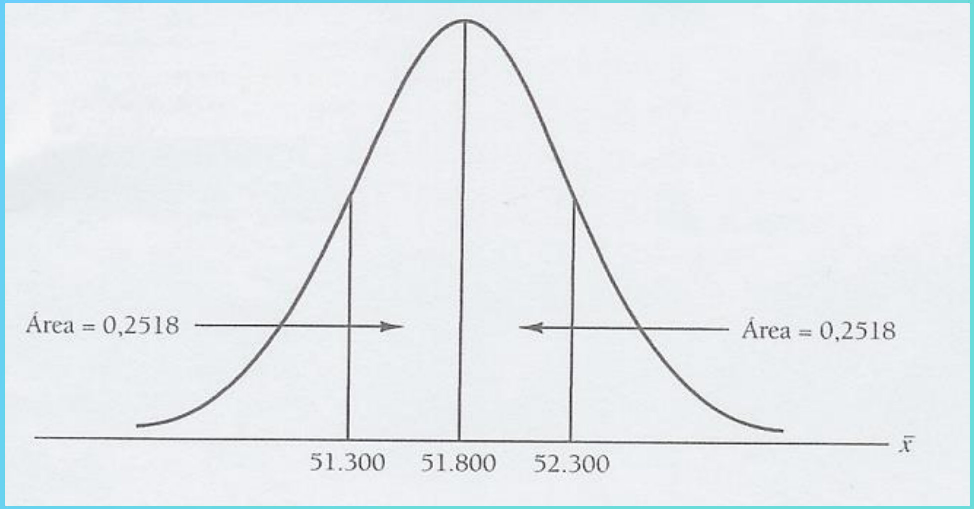


DISTRIBUIÇÃO EMPÍRICA



A distribuição empírica assemelha-se à Normal.

MODELO TEÓRICO (NORMAL)



Por meio do modelo teórico, é possível responder às seguintes questões de negócio:

Qual a probabilidade do salário anual estar entre R$ 51.300,00 e R$ 52.300,00 ?

Qual a probabilidade do salário anual ser inferior a R$ 51.300 ?

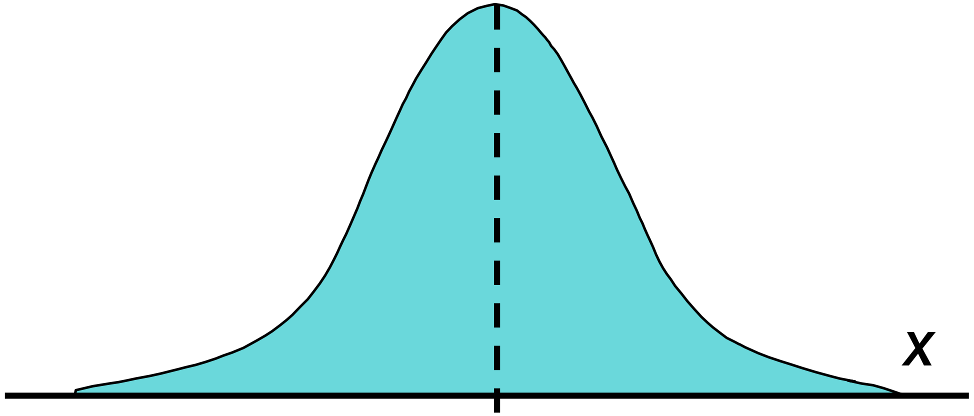
Qual a probabilidade do salário anual ser superior a R$ 51.300 ?

Distribuição Normal

A Distribuição Normal é a distribuição mais importante da Estatística.

Também é denominada Distribuição de Gauss ou Distribuição Gaussiana, em homenagem a Carl Firedrich Gauss, que assumiu que erros de medida poderiam ser modelados pela Distribuição Normal .

Carl Friedrich Gauss
(Alemanha, 1777-1855)



μ
Define o ponto central da distribuição.

Características da Normal:

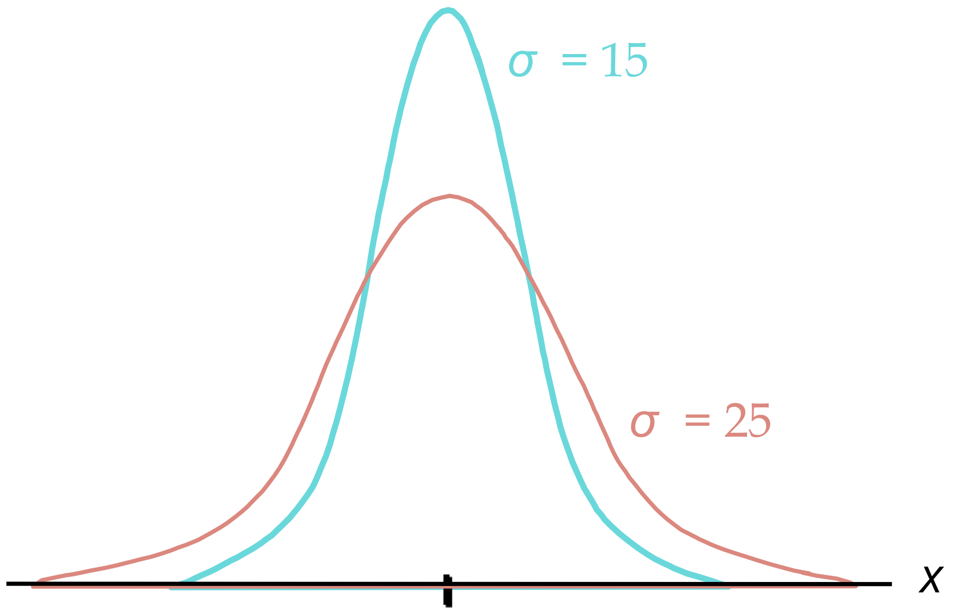
- Formato de sino
- Simétrico ao redor da média (=mediana e = moda)
- Assume valores nos reais
- Notação: $X \sim N\,(\mu, \sigma^2)$

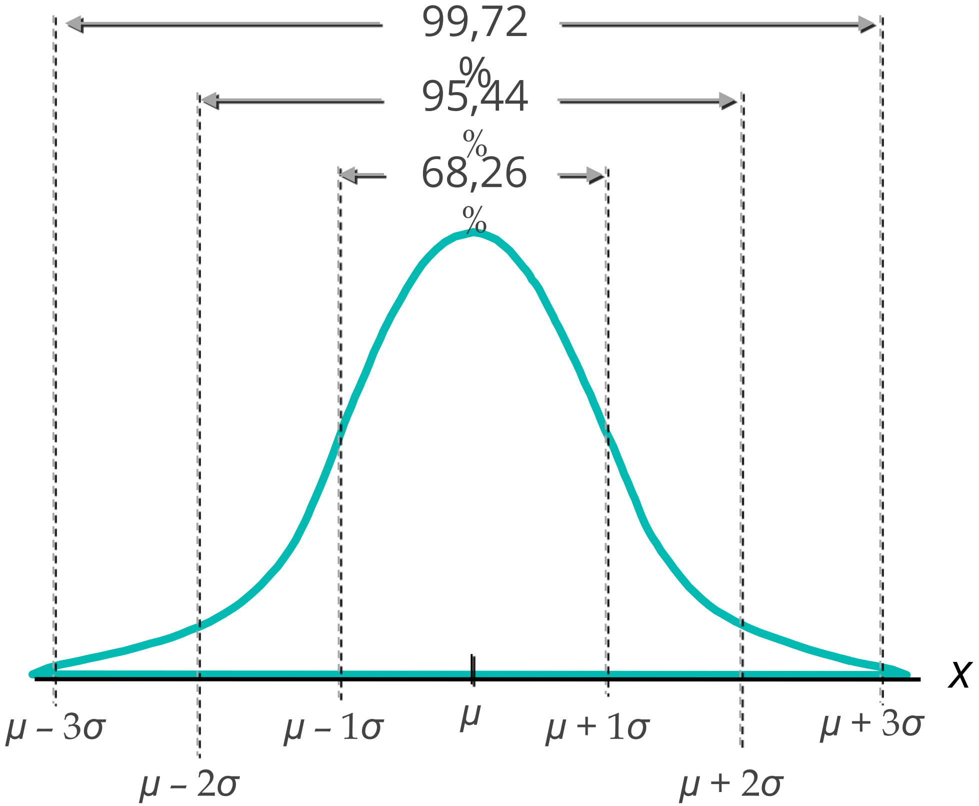
média

variância

Distribuição Normal

O desvio padrão σ define o "achatamento" da curva.





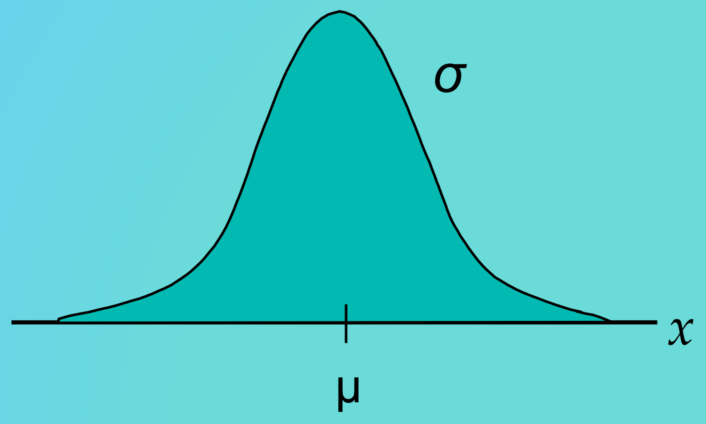
Quando uma variável segue a distribuição normal é esperado a seguinte distribuição ao redor da média μ.

Distribuição Normal Padrão

Para converter qualquer variável aleatória com distribuição Normal de média μ e variância σ^2 na **distribuição Normal Padrão** é necessário fazer a seguinte transformação:

X : variável aleatória com distribuição Normal de média μ e variância σ^2

$X \sim N(\mu, \sigma^2)$



Padronização Z-score:

$$z = \frac{(x-\mu)}{\sigma}$$

Z : variável aleatória com distribuição Normal de média μ=0 e variância $\sigma^2=1$

$Z \sim N(0, 1)$

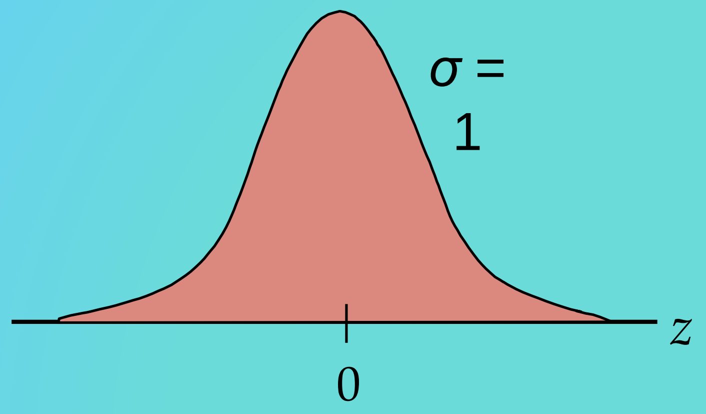
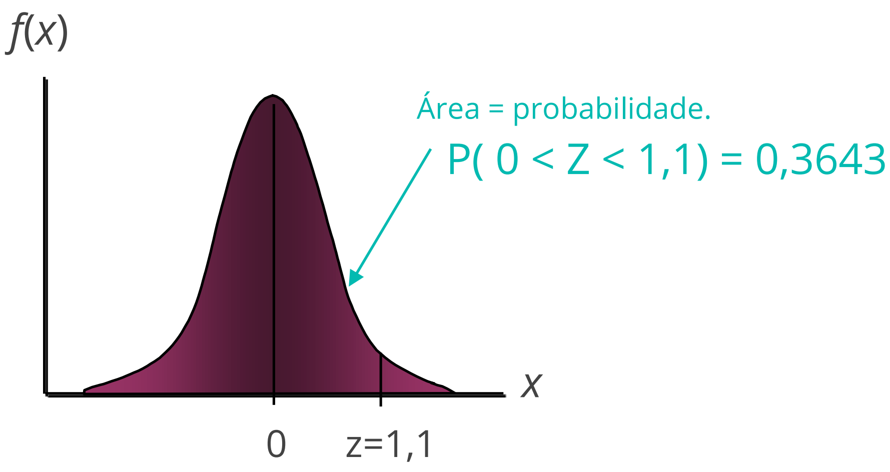
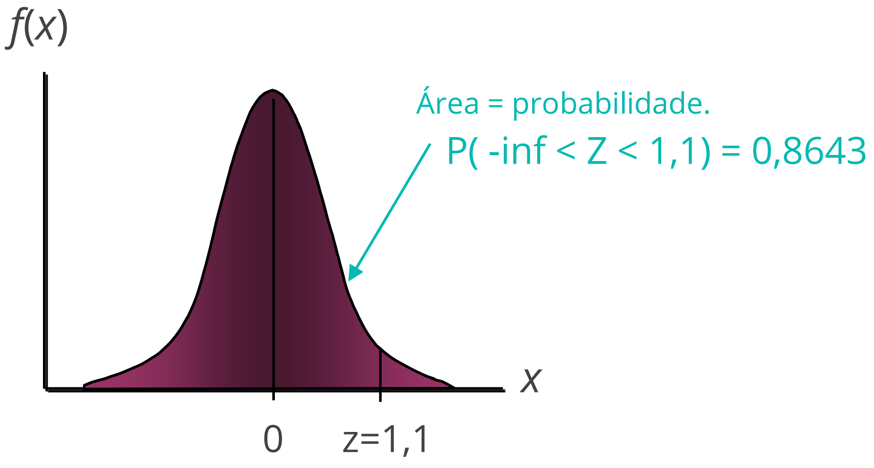


Tabela da Distribuição Normal Padrão



A área abaixo da curva entre um intervalo de [0; z] pode ser obtida consultando a tabela ao lado.

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



No Excel, a área abaixo da curva entre um intervalo de [-Inf; z] pode ser obtida por meio da função:

F(x): DIST.NORMP(z)

Passos para encontrar a probabilidade de uma variável aleatória que segue uma distribuição Normal:

1. Realizar a padronização z-score da variável X (original para Z padronizada).
2. Identificar o intervalo dos valores de z-score de interesse.
3. Consultar os valores da área abaixo da curva do intervalo de interesse:
    i. Tabela slide 37.
    ii. Utilizando a função do Excel DIST.NORMP( )

Note que tanto a Tabela (slide 37) como a função do Excel utilizam intervalos pré-definidos, e nem sempre o cálculo da probabilidade sairá direto da consulta das áreas de probabilidade.

Exemplo: Se usar o Excel, a área abaixo da curva será 0,8643, porém se o interesse for em valores entre [0; 1,1], deverá subtrair-se 0,5, obtendo-se:

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

Exemplo: Empresa de autopeças

1.ii.a. NORMAL | DISTRIBUIÇÕES DE PROBABILIDADES

Uma empresa que vende autopeças, acessórios e realiza troca de óleo do motor de carros, identificou que o volume de óleo trocado segue uma distribuição normal com média de 15 litros e desvio padrão de 6 litros.

(a) Quando o estoque do óleo chega em 20 litros, é solicitado um pedido de reposição. O gerente gostaria de saber a probabilidade de atendimento do próximo cliente, caso deixe o estoque acima de 20 litros, ou seja, P (X <20).

(b) Qual deve ser o ponto de reabastecimento, caso o gerente queira garantir que a probabilidade de não atendimento do próximo cliente esteja em 0,05?

Item (a): Empresa de autopeças

1.ii.a. NORMAL | DISTRIBUIÇÕES DE PROBABILIDADES

Passo 1: Converter *x* para distribuição z:

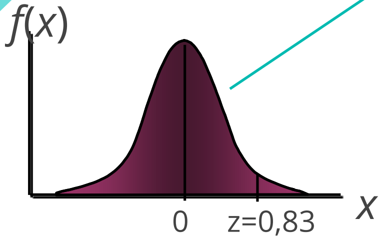
$$z = (x - \mu)/\sigma$$
$$= (20 - 15)/6$$
$$= 0{,}83$$

Passo 2: Encontre a área abaixo da curva no ponto
z = 0,83.

Passo 3: Consultar a tabela.
P(-inf < Z < 0,83) =
P(-inf < Z < 0) + P(0< Z < 0,83)=
0,5+0,2967

Passo 4: obter:
P(Z > 0,83) = 1 – 0,7967
=0,2033

P(*X* > *20*)

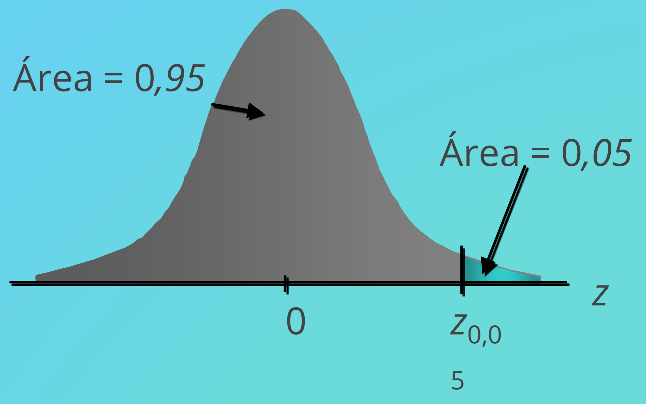


z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	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

Conclusão de negócio: A probabilidade de não atendimento do próximo cliente é de 0,2033.

Item (b): Empresa de autopeças

Passo 1: Encontre o valor de z correspondente a área de 0,05 na cauda à direita da curva.



Passo 2: Consultar na tabela $z_{,05} = 1,64$.

Passo 3: Voltar a escala original X.

$$x = \mu + z_{,05}\sigma$$
$$= 15 + 1,645 * 6$$
$$= 24,87$$

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	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0,2549
0,7	0,2580	0,2612	0,2642	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
0,8	0,2881	0,2910	0,2939	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
0,9	0,3159	0,3186	0,3212	0,3238	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	[illegible]	[illegible]	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	[illegible]	[illegible]	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	[illegible]	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	[illegible]	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	[illegible]	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

Nós olhamos o complemento da área da cauda (*1 - 0,05 = 0,95*)

Conclusão de negócio: Ao elevar o ponto de reabastecimento de 20 litros para 25, a probabilidade de não atendimento do próximo cliente diminui de 20% para 5%. Esta é uma redução significativa, que ajudara a empresa a não ficar sem estoque e, portanto, não deixar de atender seus clientes.

Considerando a variável aleatória Z com Distribuição Normal Padrão, calcule as probabilidades (4 casas decimais):

(a) $P(0 < Z < 0{,}85)$

(b) $P(0 < Z < 1{,}35)$

(c) $P(Z < -1{,}15)$

(d) $P(-0{,}82 < Z < 1{,}39)$

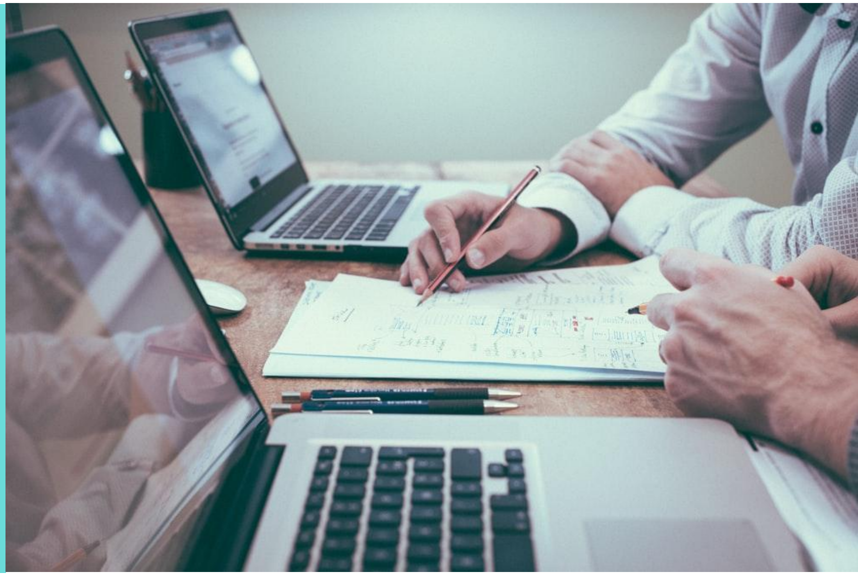
Exercício: Faturamento de empresa de consultoria

Uma empresa de consultoria possui um faturamento mensal médio de 40 milhões de reais com desvio padrão de 10 milhões de reais. Assumindo que o faturamento desta empresa segue uma distribuição normal, qual a probabilidade de que em um determinado mês o faturamento:

(a) seja superior a 50 milhões de reais ?

(b) esteja entre 30 e 50 milhões de reais?

(c) seja inferior a 30 milhões de reais?



Exercício: Vistoria

Uma seguradora possui um gasto com vistoria de automóvel médio de 20 reais com um desvio padrão de 5 reais. Calcule as probabilidades de que, em uma determinada vistoria, o gasto seja:

(a) superior a 30 reais.

(b) inferior a 18 reais.

(c) entre 15 e 28 reais.

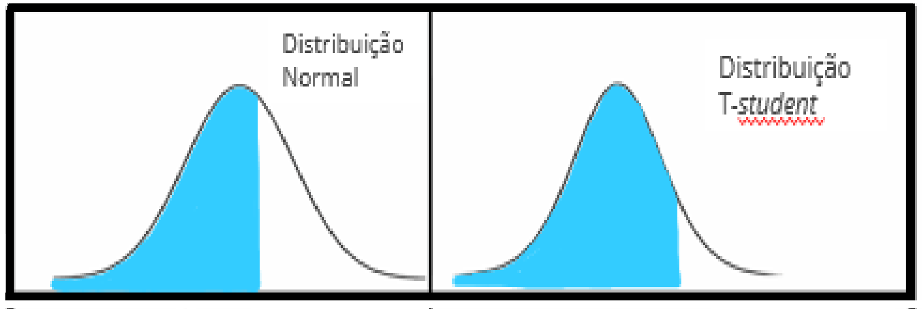
Outra Distribuição Contínua Importante

Da mesma forma que obtivemos as áreas abaixo da curva da Normal, é possível obter os valores das áreas abaixo da curva da distribuição:

- T-*student*

Cada distribuição é utilizada em casos específicos para solução de problemas práticos, e possui sua própria tabela e distribuição de probabilidades.

Outra Distribuição Contínua Importante



Todas as distribuições apresentadas, serão avaliadas até o valor crítico acumulado (área em azul da figura ao lado) para obtenção da probabilidade no Excel.

Os intervalos de valores de X para as distribuições:

- Normal e t-*student*: reais]-inf; +inf[

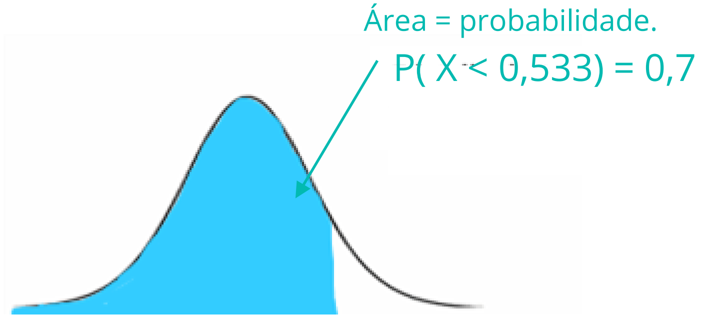
Distribuição	Nomenclatura	Parâmetros	Função do Excel	Exemplo de uso prático
Normal	N (μ, σ^2)	μ (média) e σ^2 (variância)	=DIST.NORMP.N(Z; valor acumulado até o ponto Z)	Testar a média
T-*student*	T (k)	k graus de liberdade	=DIST.T(x; k, valor acumulado até o ponto x)	Testar a média (dados amostrais)

Considerando a variável aleatória X com Distribuição T-*student* com n=20 observações, calcule as probabilidades (4 casas decimais):

(a) P (X < 0,127)

(b) P (X < -0,257)

(c) P (X > -0,391)



No Excel, a área abaixo da curva entre um intervalo de [-Inf; z] pode ser obtida por meio da função:

F(x): = DIST.T(0,533; n-1; VERDADEIRO)

Arquivo “Distribuicoes.xlsx”, aba “T-student”

2. Inferência Estatística

Análise Exploratória:

Dados observados, geralmente por meio de uma amostra

Afirmação sobre a **amostra**: O salário de gerentes de São Paulo é maior que em outros estados.

Modelos Probabilísticos:

Modelos teóricos capazes de representar comportamentos das variáveis

Qual a **distribuição teórica** adequada para inferir sobre o salário dos gerentes?

Análise Inferencial:

Argumentos estatísticos para fazer afirmações sobre a população

É esperado que na **população** de gerentes de São Paulo o salário seja maior do que em outros estados?

Objetivo da Inferência Estatística

Exemplo: Altura dos brasileiros

Parece razoável assumir que a altura dos homens do Brasil segue uma distribuição Normal.

A afirmação sobre a distribuição não é suficiente se não for conhecido os valores dos seus parâmetros, por exemplo, média e variância.

O propósito de uma pesquisa seria descobrir (estimar) os parâmetros da distribuição a partir de uma amostra.

Se pudéssemos, medir a altura de todos os brasileiros adultos, teríamos meios de obter a distribuição exata da altura e encontrar os correspondentes parâmetros.

Como os dados são provenientes de uma <u>amostra</u>, precisamos estimar os valores dos parâmetros e posteriormente inferir estes valores para população.

Importância da Inferência Estatística

2.i. CONCEITOS PRELIMINARES | APLICAÇÕES REAIS

Teste de durabilidade da lâmpada. Não faz sentido testar todas as lâmpadas até elas queimarem

Teste de lançamento de um novo remédio, quando já existe uma população com uso do tratamento tradicional

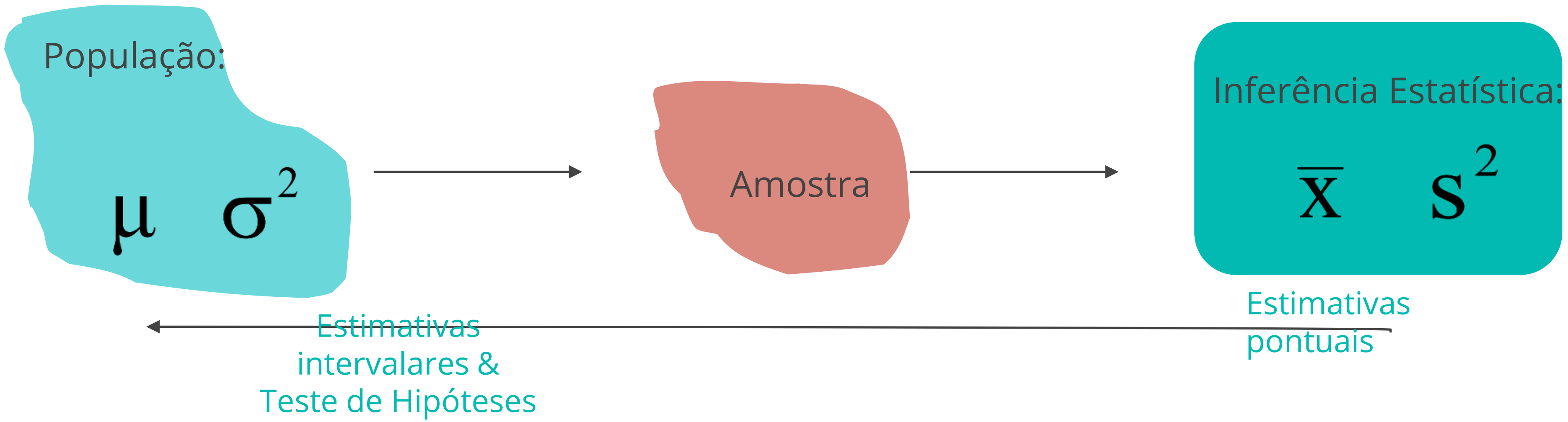
Gôndola do supermercado: Existe uma evidência de igualdade de médias de vendas para as 4 alturas de gôndolas

Soluções para casos apresentados são objetivos da inferência estatística.
Conceitos:

- ✓ **População:** conjunto de todos os elementos
- ✓ **Amostra:** qualquer subconjunto da população

Objetivo da Inferência Estatística

O objetivo da Inferência Estatística é produzir afirmações sobre alguma característica da população, na qual estamos interessados, a partir de informações colhidas (amostras) de uma parte dessa população.



Distribuição amostral da média

Consideremos uma população identificada pela variável X, cujo os parâmetros são:

- média populacional $\mu=E(X)$
- variância populacional $\sigma^2=Var(X)$

Considere uma amostra aleatória de tamanho n de uma determinada população. Seja $\bar{X}$ a média amostral, temos que:

$E(\bar{X}) = \mu$

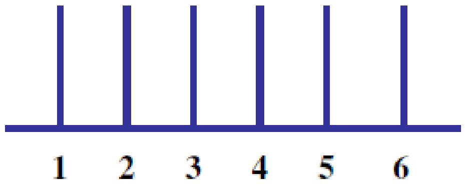
$Var(\bar{X}): \sigma^2/n$

Logo, a média e a variância de $\bar{X}$ são dadas por μ e σ^2/n, respectivamente.

Exemplo do lançamento de dados

Lançamento de 1 dado:

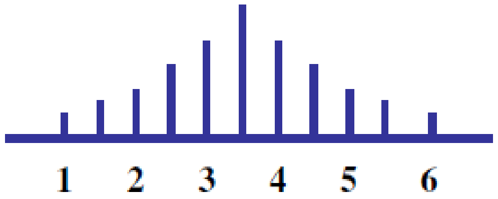
$\bar{X}$	1	2	3	4	5	6
$P(\bar{X})$	1/6	1/6	1/6	1/6	1/6	1/6



Distribuição da média das faces

Lançamento de 2 dados:

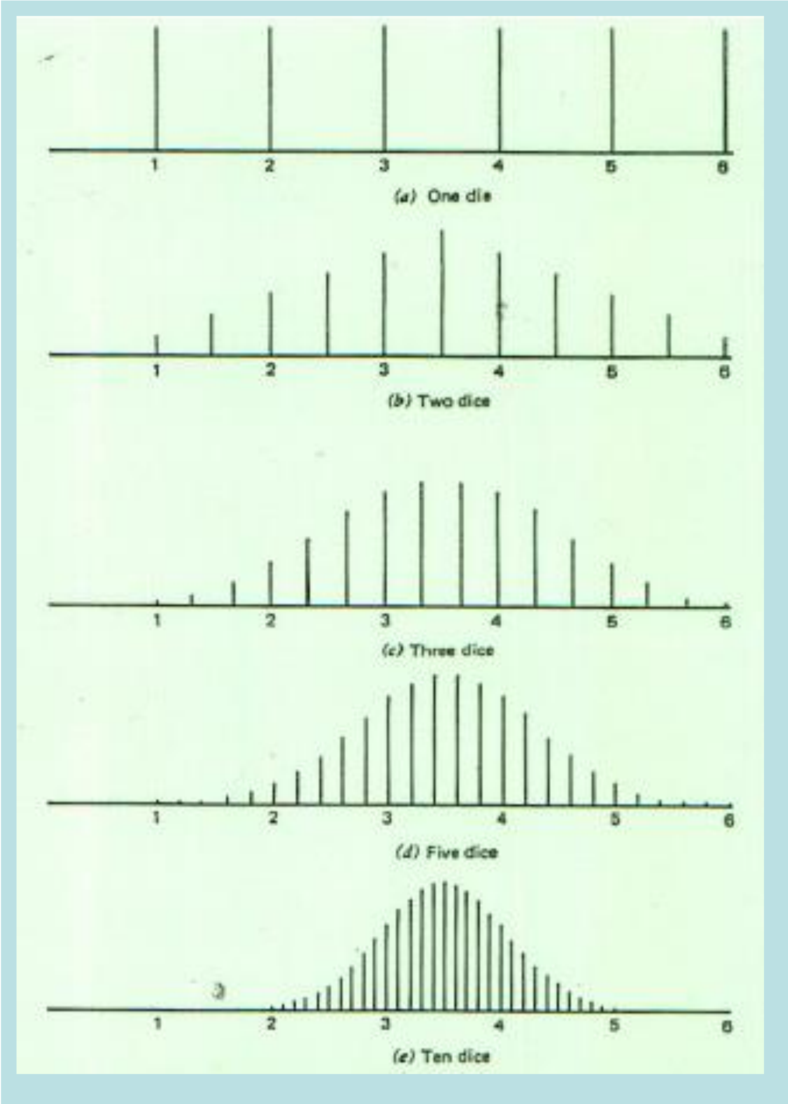
S	1	2	3	4	5	6
1	1 1	1 2	1 3	1 4	1 5	1 6
2	2 1	2 2	2 3	2 4	2 5	2 6
3	3 1	3 2	3 3	3 4	3 5	3 6
4	4 1	4 2	4 3	4 4	4 5	4 6
5	5 1	5 2	5 3	5 4	5 5	5 6
6	6 1	6 2	6 3	6 4	6 5	6 6



Distribuição da média das faces

$\bar{X}$	1	1,5	2	2,5	3	3,5	4	4,5	5	5,5	6	
$P(\bar{X})$	1	2	3	4	5	6	5	4	3	2	1	/ 36

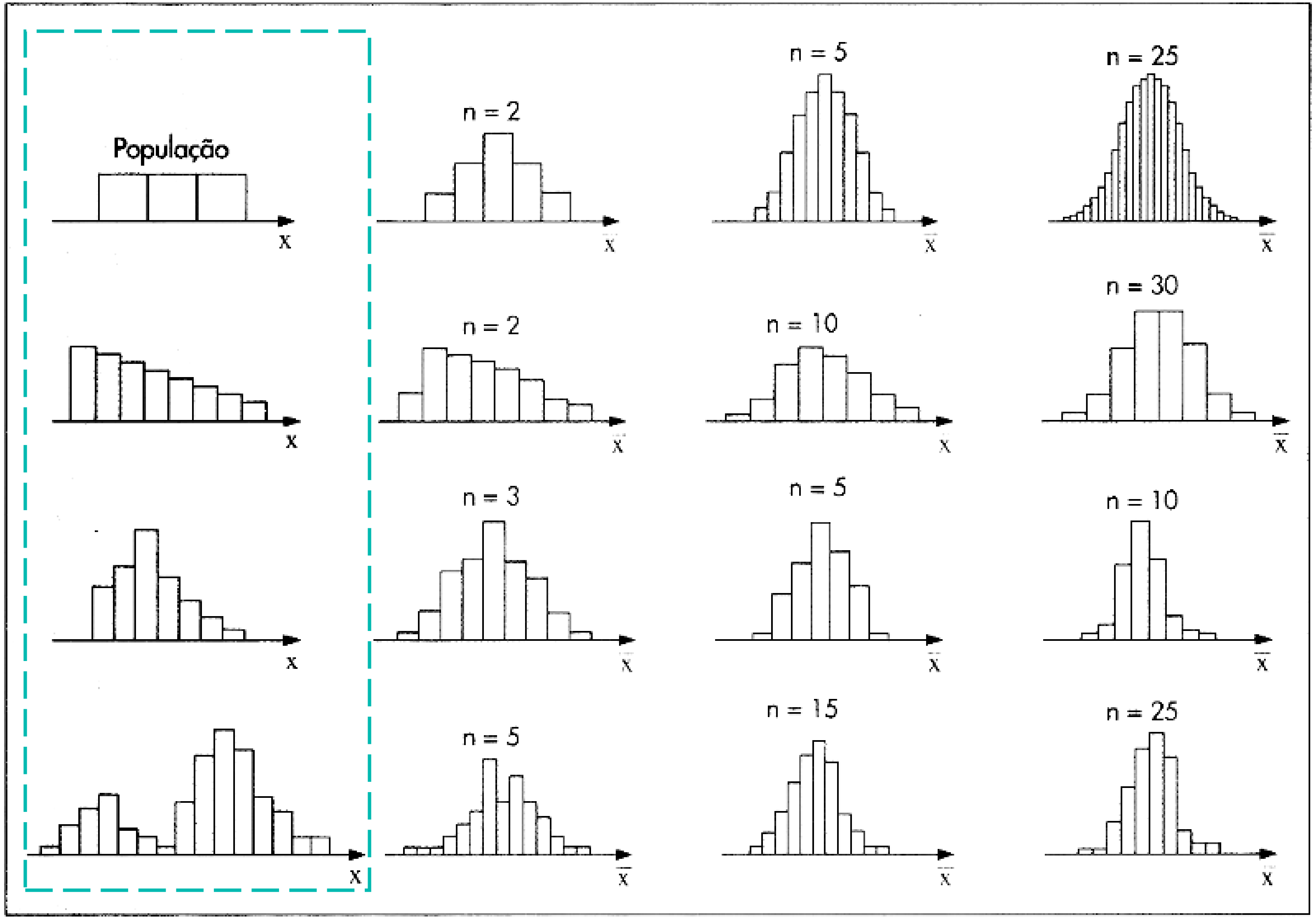
Exemplo do lançamento de dados



Distribuição amostral da média:

$$\overline{X} \overset{n \to \infty}{\sim} N(\mu\ ;\ \sigma^2/n)$$

Exemplo outras distribuições



Pela figura, note que conforme *n* vai aumentando, independente da forma da distribuição da população, a distribuição amostral da média aproxima-se cada vez mais da **distribuição Normal**.

Este resultado fundamental da teoria da Inferência Estatística é conhecido como Teorema do Limite Central (TLC).

TLC - Teorema do Limite Central

Distribuição amostral da média:

$$\overline{X} \overset{n \to \infty}{\sim} N(\mu \; ; \sigma^2 / n)$$

Teorema do Limite Central (TLC):

Para *n* grande (*n* > *30*), proveniente de uma amostra aleatória simples, o TLC nos permite concluir que a distribuição amostral de $\bar{X}$ pode ser aproximada por uma distribuição normal.

Para *n* pequeno (*n* < *30*) a distribuição amostral de $\bar{X}$ pode ser considerada normal somente se a população tiver distribuição normal.

Onde utilizaremos estes conceitos de inferência estatística?

Utilizamos a inferência estatística para várias aplicações, por exemplo:

1. Estimação de parâmetros populacionais por meio de INTERVALO DE CONFIANÇA.
2. TESTE DE HIPÓTESES para tomada de decisão.

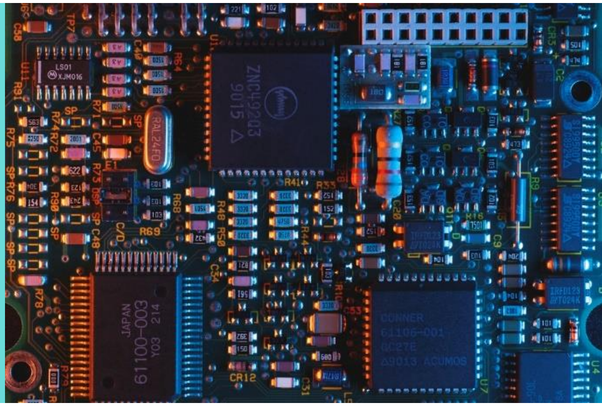
Serão temas de estudo das próximas aulas....

3. Exercícios

3.i. Funcionamento do sistema

EXERCÍCIOS - CASE

No sistema abaixo, cada componente tem probabilidade 0,7 de funcionar. Se uma das componentes falhar, o sistema não funciona. Suponha independência de funcionamento das componentes, qual a probabilidade de:



(a) O sistema funcionar?

(b) O sistema não funcionar?

(c) Exatamente 2 componentes funcionarem?

Calcular as funções de probabilidades no Excel.

3.ii. Faturamento de empresa de consultoria

EXERCÍCIOS - CASE

Estudos meteorológicos indicam que a precipitação pluviométrica mensal, em períodos de seca numa certa região, pode ser considerada seguindo uma distribuição Normal de média 30 mm e variância 16 mm^2.

(a) Qual seria o valor da precipitação pluviométrica de modo que exista apenas 10% de probabilidade de haver uma precipitação inferior a esse valor?

(b) Construa um intervalo central em torno da média que contenha 80% dos possíveis valores de precipitação pluviométrica.

(c) Admitindo esse modelo correto para os próximos 50 meses, em quantos deles esperaríamos uma precipitação pluviométrica superior a 34 mm?

Calcular as funções de probabilidades no Excel ou consultar tabela da Normal.

Referência

LIVRO-TEXTO

1. Anderson, R. A., Sweeney, J. D. e Williams, T. A. *Estatística Aplicada à Administração e Economia.* Editora Cengage. 4ª edição, 2019.