Preditiva.ai

Probabilidades

Variáveis Aleatórias

Probabilidades

Variáveis Aleatórias: Introdução

Em **Estatística Descritiva** aprendemos que as tabelas de frequência são úteis para resumir uma variável de um conjunto de dados. Vejamos abaixo a tabela que resume a variável "**Quantidade de Celulares em uma Residência**":

Qte de Celulares	0	1	2	3	4	Total
Frequência Relativa ou Probabilidade	10%	40%	30%	15%	5%	**100%**

Variáveis cuja probabilidade de ocorrência tem **caráter aleatório** são chamadas de **Variáveis Aleatórias**.

E existem dois tipos de **Variáveis Aleatórias**:

- Variáveis Aleatórias **Discretas ou Qualitativas**
- Variáveis Aleatórias **Contínuas**

Probabilidades
Variáveis Aleatórias Discretas ou Qualitativas

Probabilidades

Variáveis Aleatórias Discretas ou Qualitativas

Quando estudamos uma variável aleatória discreta ou qualitativa, podemos representa-la no formato abaixo:

Qte de Celulares	**0**	**1**	**2**	**3**	**4**
Probabilidade	10%	40%	30%	15%	5%

Desta forma, se quisermos calcular a probabilidade **P (Qte de Celulares = 3)** basta olhar na tabela e buscar a probabilidade associada ao valor de interesse.

Portanto, **P (Qte de Celulares = 3) = 15%.**

A função acima, que **atribui uma probabilidade** a cada valor da variável aleatória é chamada de **Função de Probabilidade**.

A notação dessa função é dada por: $P(X = x_i) = p(x_i) = p_i$ ou

X	**x_1**	**x_2**	**x_3**
P_i	p_1	p_2	p_3

(Formato algébrico) (Formato tabular)

Probabilidades

Variáveis Aleatórias Discretas ou Qualitativas

Da mesma forma que calculamos a frequência acumulada nas tabelas, podemos também calcular a **probabilidade acumulada** de uma variável aleatória. Veja a seguir:

Qte de Celulares	**0**	**1**	**2**	**3**	**4**
Probabilidade	10%	40%	30%	15%	5%
Prob. Acum.	10%	50%	80%	95%	100%

Se quisermos calcular a probabilidade **P (Qte de Celulares $\leq$ 2)** , ou seja, a probabilidade de uma residência ter até 2 celulares, basta consultar a Probabilidade Acumulada associada ao valor 2.

Ou seja, **P (Qte de Celulares $\leq$ 2) = P (Qte = 0) + P (Qte = 1) + P (Qte = 2) = 10% + 40% + 30%** = 80%

A função acima, que **acumula as probabilidades** até um dado valor da variável aleatória é chamada **Função de Distribuição de Probabilidade**.

A notação dessa função é dada por: $F(x) = P(X \leq x)$

Probabilidades

Variáveis Aleatórias Discretas ou Qualitativas

Qte de Celulares	0	1	2	3	4
Probabilidade	10%	40%	30%	15%	5%
Prob. Acum.	10%	50%	80%	95%	100%

Assim como calculamos a probabilidade **P (Qte de Celulares $\leq$ 2)** , podemos calcular **P (Qte de Celulares $>$ 2)**, que é a probabilidade da quantidade ser maior que 2.

Podemos calcular de duas maneiras:

- **P (Qte de Celulares $>$ 2) = P (Qte = 3) + P (Qte = 4) = 15% + 5% = 20%**

Ou podemos calcular sabendo que a probabilidade de todos os eventos é de 100% e que o **evento complementar** é **P (Qte de Celulares $\leq$ 2)** que calculamos anteriormente:

Dois eventos A e B são complementares se:

- P (A $\cup$ B) = 100% e
- P (A $\cap$ B) = 0%.

- **P (Qte de Celulares $>$ 2)** = 100% - P (Qte de Celulares $\leq$ 2) = 100% - 80% **= 20%**

Revisão

Vimos que as **Variáveis Aleatórias Discretas ou Qualitativas** devem ser utilizadas quando a variável de interesse é discreta ou qualitativa.

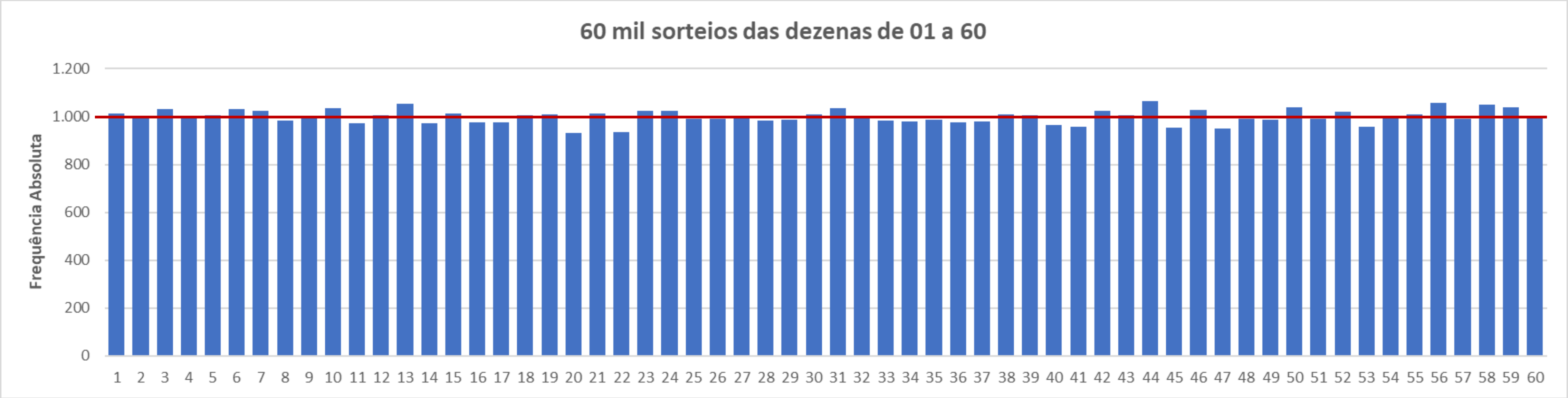
Além disso, vimos como podemos calcular a **probabilidade** de ocorrência de um **valor específico**, e da **probabilidade acumulada**.

Probabilidades

Distribuição de Probabilidades Uniforme

Probabilidades
Variáveis Aleatórias Discretas: Uniforme

Auditores foram chamados para auferir as 60 esferas que fazem parte do sorteio da Mega Sena. Assim, desenvolveram um dispositivo que realizou 60 mil sorteios. O resultado está representado abaixo:



Com base nesse gráfico de frequência absoluta, você acha que a **probabilidade** de alguma das dezenas ser sorteada é **maior** do que de outras?

Quando os resultados de um **fenômeno de caráter aleatório** possuem a **mesma probabilidade de ocorrência**, dizemos que esse fenômeno tem **Distribuição Uniforme**.

Probabilidades

Variáveis Aleatórias Discretas: Uniforme

Preditiva.ai

Uma variável aleatória **X** segue a distribuição **Uniforme** se a sua função de probabilidade atribuir a mesma probabilidade **1/k** para para um dos **k** valores da variável. Ou seja, sua função é dada por:

$$P(X = x_i) = \frac{1}{k}, \text{sendo } i = 1, 2, 3 \ldots, k.$$

A notação utilizada será $X \sim U(a, b)$, sendo ***a*** e ***b*** o **menor** e o **maior** valor que a variável pode assumir, respectivamente.

ALEATÓRIO(): gera um número entre 0 e 1.

ALEATÓRIOENTRE(a ; b): gera um número inteiro entre **a** e **b**.

Revisão

Vimos que a **Distribuição Uniforme** é utilizada quando todos os resultados de uma variável aleatória possuem exatamente a **mesma probabilidade de ocorrerem**.

Preditiva.ai

Probabilidades
Distribuição de Probabilidades Binomial

O que você verá nessa aula?

- ❑ Introdução a Variáveis Aleatórias
- ❑ Variáveis Aleatórias Discretas ou Qualitativas
  - ❑ Distribuição Uniforme
  - ❑ Distribuição Binomial
- ❑ Variáveis Aleatórias Contínuas
  - ❑ Distribuição Normal
- ❑ Outras Distribuições de Probabilidades

Probabilidades

Variáveis Aleatórias Discretas: Binomial

Preditiva.ai

Um vendedor recebeu uma lista de **20 potenciais clientes** para oferecer seus produtos. Sabendo que a **probabilidade de vender** para cada um deles é de **5%**, qual é a probabilidade de que ele consiga vender para exatamente 2 clientes?

Combinação

Prob. Vendas: $(5\%)^2$

Prob. Não Vendas: $(95\%)^{18}$

95%	5%	95%	95%	95%	95%	95%	95%	95%	95%	95%	5%	95%	95%	95%	95%	95%	95%	95%	95%

Combinação

Prob. Vendas: $(5\%)^2$

Prob. Não Vendas: $(95\%)^{18}$

95%	95%	95%	5%	95%	95%	95%	95%	95%	95%	95%	5%	95%	95%	95%	95%	95%	95%	95%	95%

$$P(X = 2) = \binom{20}{2} 5\%^2 (1 - 5\%)^{20-2} = 18{,}9\%$$

DISTR.BINOM(2;20;5%;FALSO) = 18,9%

Probabilidades

Variáveis Aleatórias Discretas: Binomial

Uma variável aleatória **X** segue a distribuição Binomial se o fenômeno que se deseja descrever é o **número de sucessos em uma sequência de tentativas**. Cada tentativa possui apenas dois resultados possíveis: **sucesso** ou **fracasso**, as **tentativas são independentes** e a **probabilidade de sucesso** se mantém **constante** em todas as tentativas.

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \text{sendo } 0 \leq k \leq n$$

A notação utilizada será

$X \sim Bin(n, p)$, sendo:

- ***n*** o número de tentativas
- ***p*** a probabilidade de sucesso
- **k** o número de sucessos

DISTR.BINOM (k ; n ; p ; FALSO): calcula a probabilidade para **X = k**

DISTR.BINOM (k ; n ; p ; VERDADEIRO): calcula a probabilidade para **X ≤ k**

Revisão

Vimos que a **Distribuição Binomial** é utilizada quando queremos calcular a probabilidade de obter **k sucessos em n tentativas**.

Probabilidades
Distribuição de Probabilidades
Poisson

Probabilidades

Variáveis Aleatórias Discretas: Poisson

Um analista recebeu a base de dados de uma campanha de Marketing realizada por sua empresa no Google Ads. A base contém a **quantidade de cliques** em um anúncio, **em intervalos de 1 hora**, ao longo do período de 12 horas. O Diretor de Marketing te pergunta então qual é a **probabilidade** de que em **2 dias** (**de 12 horas cada**) o anúncio tenha até **60 cliques**?

Hora	Cliques
8 -| 9	2
9 -| 10	5
10 -| 11	2
11 -| 12	1
12 -| 13	0
13 -| 14	3
14 -| 15	4
15 -| 16	3
16 -| 17	4
17 -| 18	2
18 -| 19	1
19 -| 20	3
Total	**30**

Cliques	Qte Períodos 1 Hora	Freq. Relativa ou Probabilidade
0	1	8,3%
1	2	16,7%
2	3	25,0%
3	3	25,0%
4	2	16,7%
5	1	8,3%
Total	**12**	**100,0%**

Informações extraídas:

- Total de Cliques em 12h: 30
- Média de cliques por hora: **2,5**

$$P(X \leq 60) = \sum_{k=0}^{60} \frac{e^{-2,5*24}(2,5*24)^k}{k!} = 53,4\%$$

DISTR.POISSON (60 ; 2,5*24 ; VERDADEIRO) = 53,4%

Probabilidades

Variáveis Aleatórias Discretas: Poisson

Uma variável aleatória **X** segue a distribuição **Poisson** com parâmetro $\lambda > 0$ se o fenômeno que se deseja descrever é a **quantidade de ocorrências de um evento** em um **intervalo de tempo determinado**.

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!} \; sendo\; k = 0,1,2,\dots.$$

A notação utilizada será

$$X \sim Po(\lambda)$$

O parâmetro λ é denominado **Taxa de Ocorrência** e representa a **média de ocorrências do evento** em um **intervalo de tempo determinado**.

DISTR.POISSON (k ; λ ; FALSO): calcula a probabilidade para **X = k**

DISTR.POISSON (k ; λ ; VERDADEIRO): calcula a probabilidade para **X ≤ k**

Revisão

Vimos que a **Distribuição Poisson** é utilizada quando queremos calcular a probabilidade de obter uma certa **quantidade de eventos** em um determinado **espaço de tempo**.

Preditiva.ai

Probabilidades

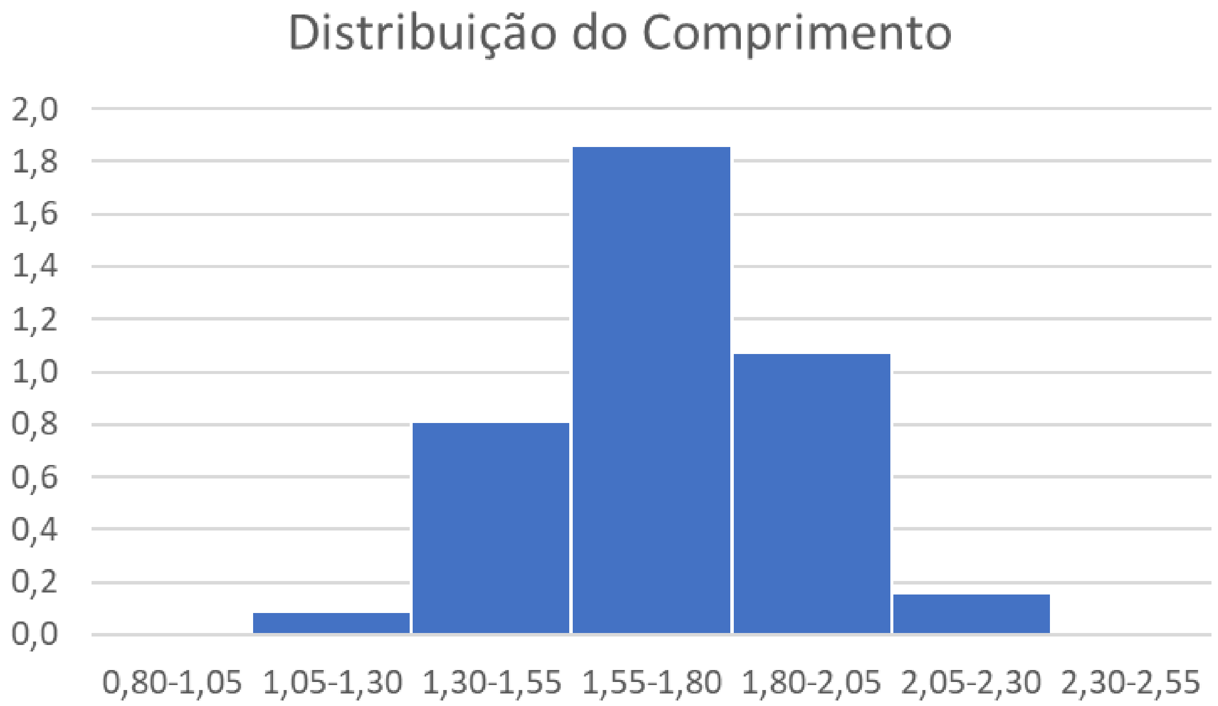
Variáveis Aleatórias Contínuas

Probabilidades
Variáveis Aleatórias Contínuas

E se tivermos uma **Variável Aleatória Contínua**? Como fazemos para calcular a probabilidade de ocorrência de um evento?

Exemplo: Uma fábrica que produz peças de plástico, para avaliar a qualidade de sua produção, realizou diversas **medições do comprimento** das peças ao longo do mês. Os resultados são apresentados abaixo:

Comprimento (cm)	Frequência relativa ou Probabilidade
0,80 a 1,05	0,1%
1,05 a 1,30	2,2%
1,30 a 1,55	20,4%
1,55 a 1,80	46,5%
1,80 a 2,05	26,9%
2,05 a 2,30	3,9%
2,30 a 2,55	0,1%
TOTAL	**100,0%**



Probabilidades
Variáveis Aleatórias Contínuas

As peças com comprimento inferior a **1,24cm** ou superior a **2,11cm** são consideradas defeituosas pois estão fora da especificação e não poderão ser aproveitadas.

Como podemos calcular a **probabilidade** de uma peça fabricada ser **defeituosa** utilizando a **tabela de frequências**?

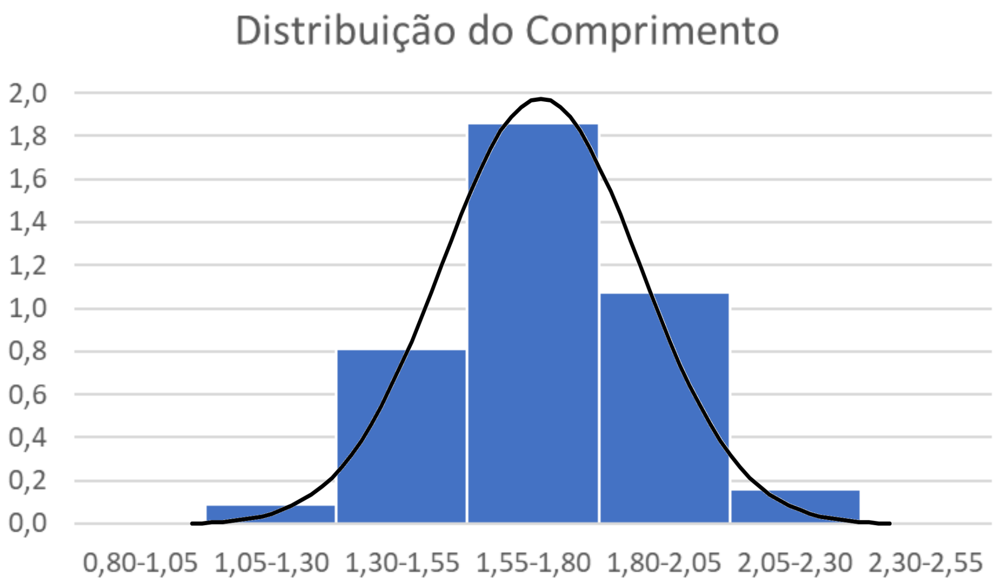
Comprimento (cm)	Frequência relativa ou Probabilidade
0,80 a 1,05	0,1%
1,05 a 1,30	2,2%
1,30 a 1,55	20,4%
1,55 a 1,80	46,5%
1,80 a 2,05	26,9%
2,05 a 2,30	3,9%
2,30 a 2,55	0,1%
TOTAL	**100,0%**

Quando analisamos uma Variável Aleatória Contínua, a tabela de frequências não é o método mais adequado, pois **ela não permite obter as probabilidades em intervalos diferentes** daqueles existentes na tabela.

Probabilidades
Variáveis Aleatórias Contínuas

As peças com comprimento inferior a **1,24cm** ou superior a **2,11cm** são consideradas defeituosas pois estão fora da especificação e não poderão ser aproveitadas.

Como podemos calcular a **probabilidade** de uma peça fabricada ser **defeituosa** utilizando a **tabela de frequências**?



Neste caso, utilizamos as **funções de probabilidades contínuas** mais adequadas para calcular as probabilidades desejadas.

Revisão

Vimos que as **Variáveis Aleatórias Contínuas** devem ser utilizadas quando a variável de interesse é contínua.

Também vimos que nesse caso, calcular as probabilidades a partir da tabela de frequências não é a melhor forma, devemos utilizar as **funções de probabilidades contínuas**, que veremos a seguir.

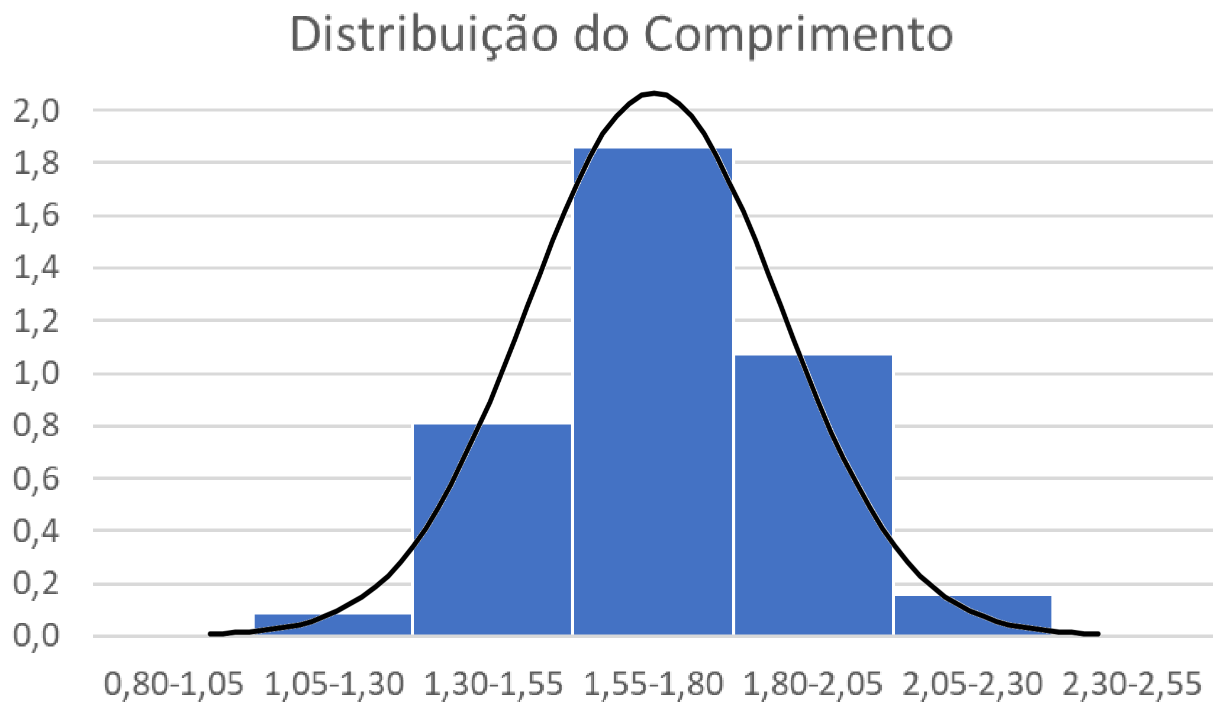
Preditiva.ai

Probabilidades
Distribuição de Probabilidades Normal

Probabilidades

Variáveis Aleatórias Contínuas: Normal

Uma fábrica que produz peças de plástico, precisa calcular a **probabilidade de uma peça fabricada ser defeituosa**. As peças com comprimento inferior a **1,24cm** ou superior a **2,11cm** são consideradas defeituosas pois estão fora da especificação e não poderão ser aproveitadas.



O que podemos extrair de **informações** deste **histograma**?

1. Distribuição simétrica
2. Concentração de valores na posição central
3. Densidade tende a zero para valores muito baixos ou valores muito altos
4. Formato de "sino"

Conclusão: A distribuição **Normal** parece ser adequada

Probabilidades

Variáveis Aleatórias Contínuas: Normal

Tendo definido a distribuição Normal como mais adequada, o 2º passo é calcular as **medidas resumo Média** e **Desvio Padrão** para podermos utilizar sua **função de probabilidade**.

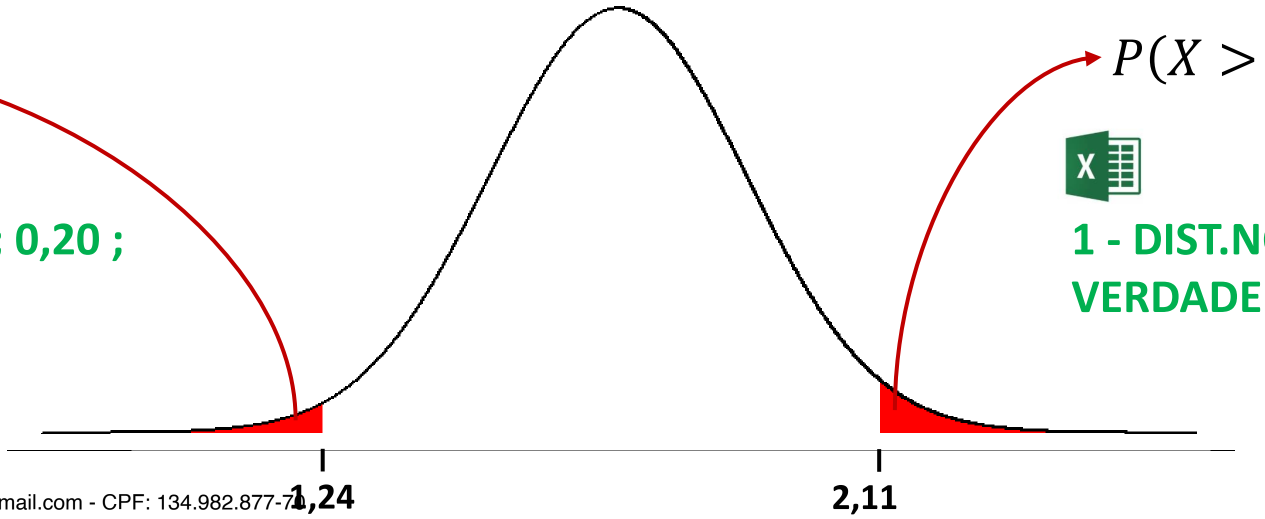
Medida Resumo	Média	Desvio Padrão
Comprimento (cm)	1,70	0,20

No 3º e último passo utilizamos a função de probabilidade para calcular a probabilidade de uma peça fabricada ser defeituosa:

$$P(\text{Defeituosa}) = P(X < 1{,}24) + P(X > 2{,}11)$$

$P(X < 1{,}24)$

DIST.NORM.N (1,24 ; 1,70 ; 0,20 ; VERDADEIRO) = 1,1%



$P(X > 2{,}11) = 1 - P(X \leq 2{,}11)$

1 - DIST.NORM.N (2,11 ; 1,70 ; 0,20 ; VERDADEIRO) = 2,0%

Probabilidades

Variáveis Aleatórias Contínuas: Normal

Logo, considerando que:

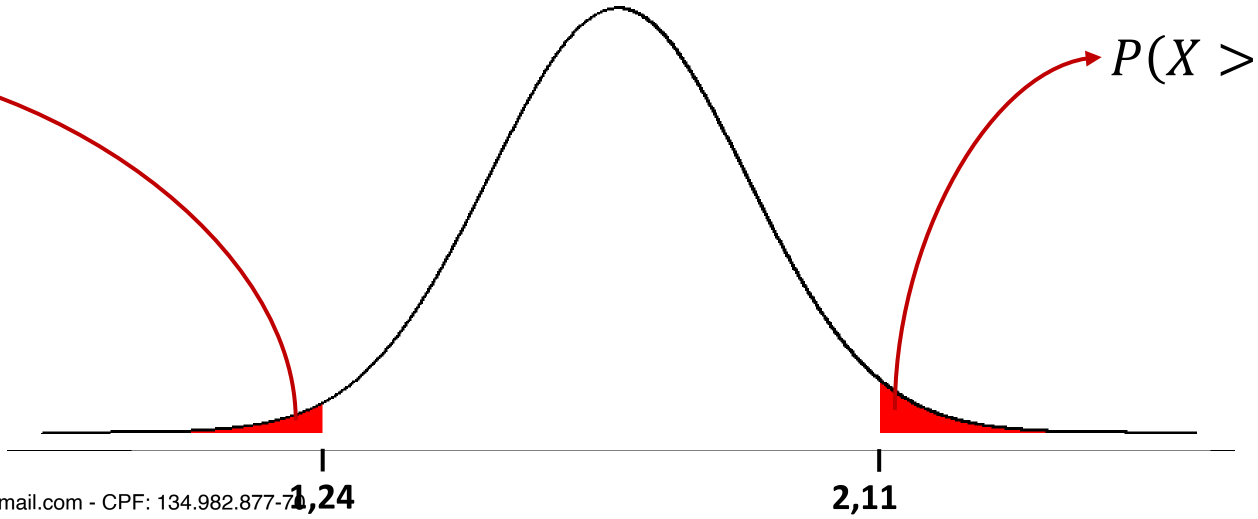
- A distribuição **Normal** é adequada
- A **média** amostral é igual a 1,70cm
- O **desvio padrão** amostral é igual a 0,20cm

A probabilidade de uma peça fabricada ser defeituosa é de **3,1%**.

$$P(\text{Defeituosa}) = P(X < 1{,}24) + P(X > 2{,}11) = 3{,}1\%$$

$P(X < 1{,}24)$=1,1%

$P(X > 2{,}11) = 2{,}0\%$



Probabilidades

Variáveis Aleatórias Contínuas: Normal

Uma variável aleatória contínua **X** segue a distribuição **Normal** com parâmetros μ e σ^2 se sua **Função Densidade de Probabilidade** é dada por:

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

A notação utilizada será $X \sim N(\mu, \sigma^2)$ sendo:

- μ a média da variável X
- σ^2 a variância da variável X

DIST.NORM.N (x ; μ ; σ ; FALSO): calcula a densidade de probabilidade para **X = x**

DIST.NORM.N (x ; μ ; σ ; VERDADEIRO): calcula a probabilidade para **X ≤ x**

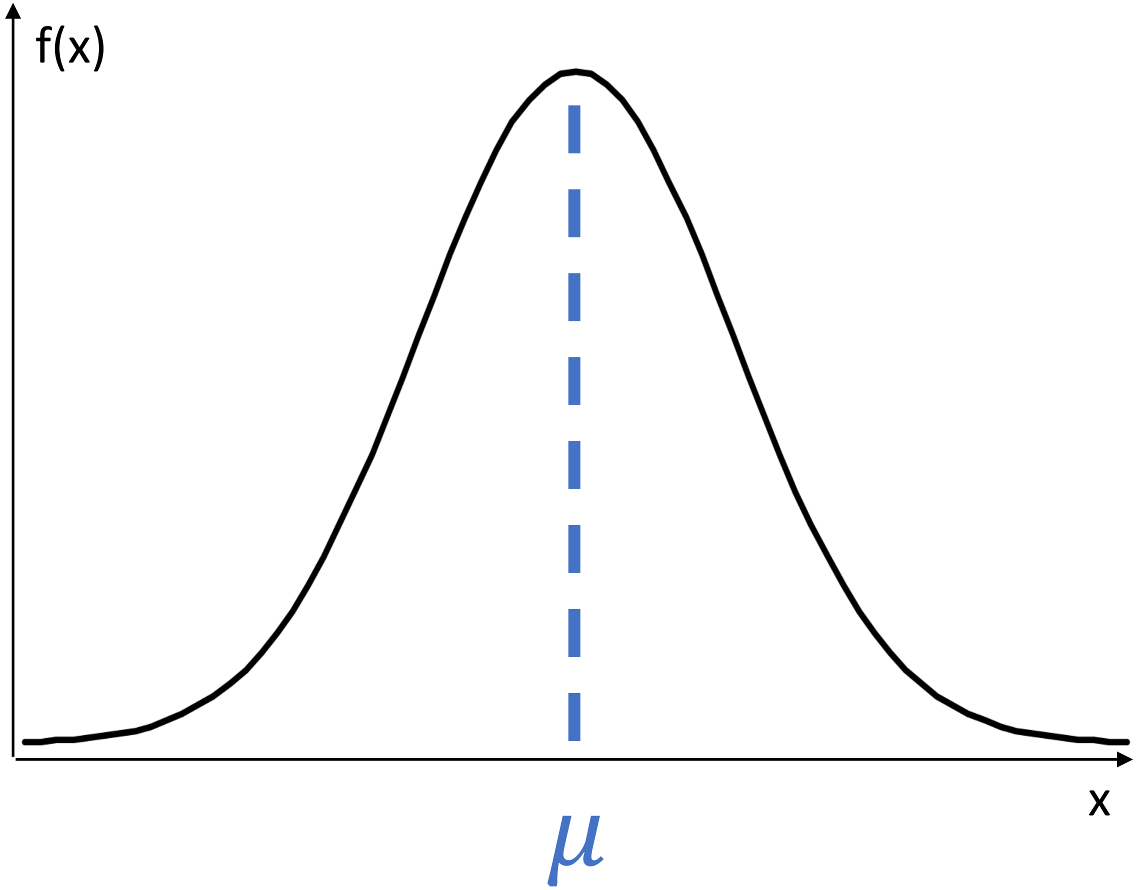
Probabilidades

Variáveis Aleatórias Contínuas: Normal

Algumas propriedades da Função de Densidade **Normal** são facilmente verificadas no seu gráfico. Veja:

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

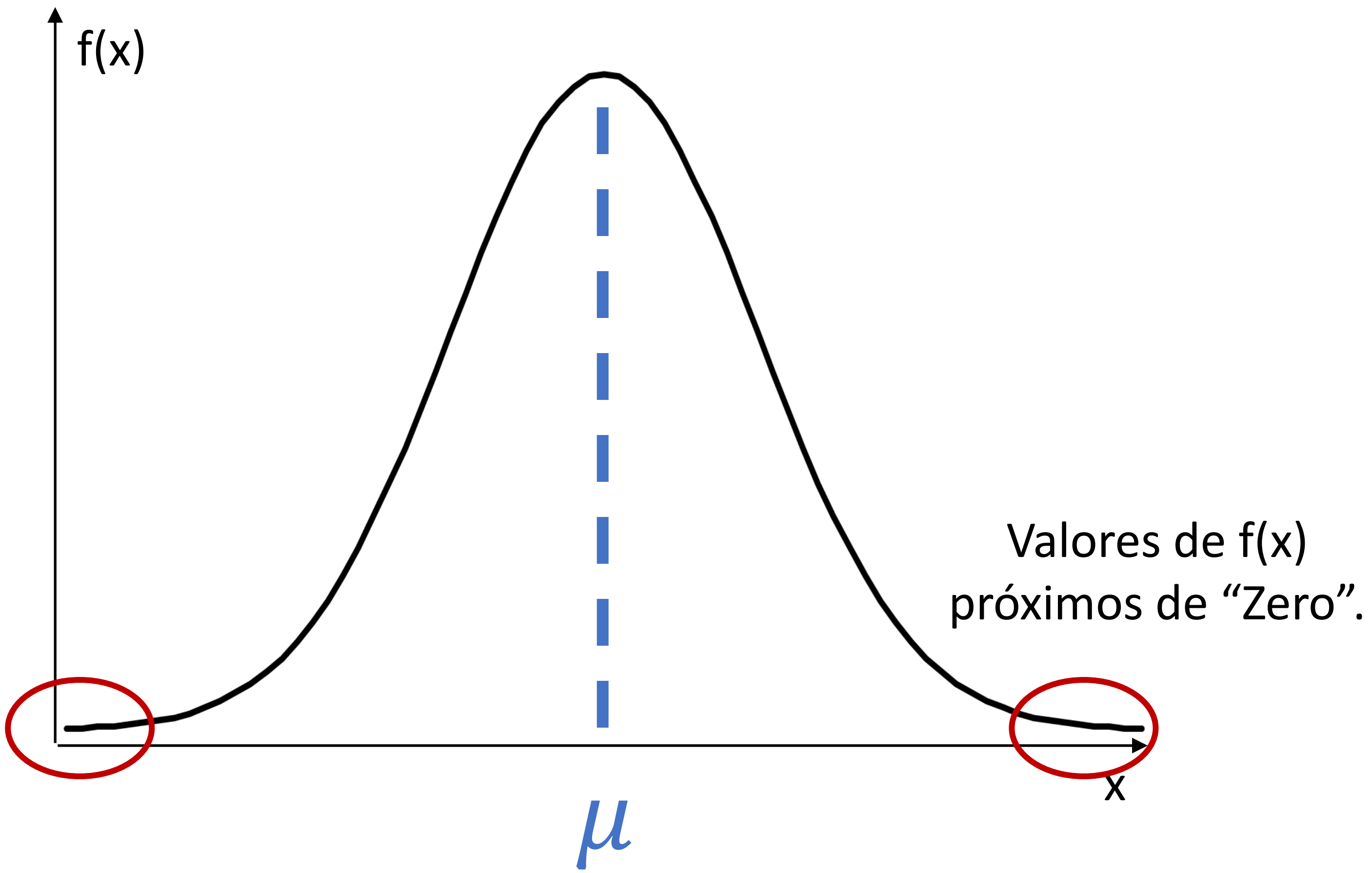
Propriedade 1: O valor máximo de f(x) é no ponto **x = $\boldsymbol{\mu}$**.



Algumas propriedades da Função de Densidade **Normal** são facilmente verificadas no seu gráfico. Veja:

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

Propriedade 2: f(x) tende a 0 quando **x** tende ao mais infinito e ao menos infinito.



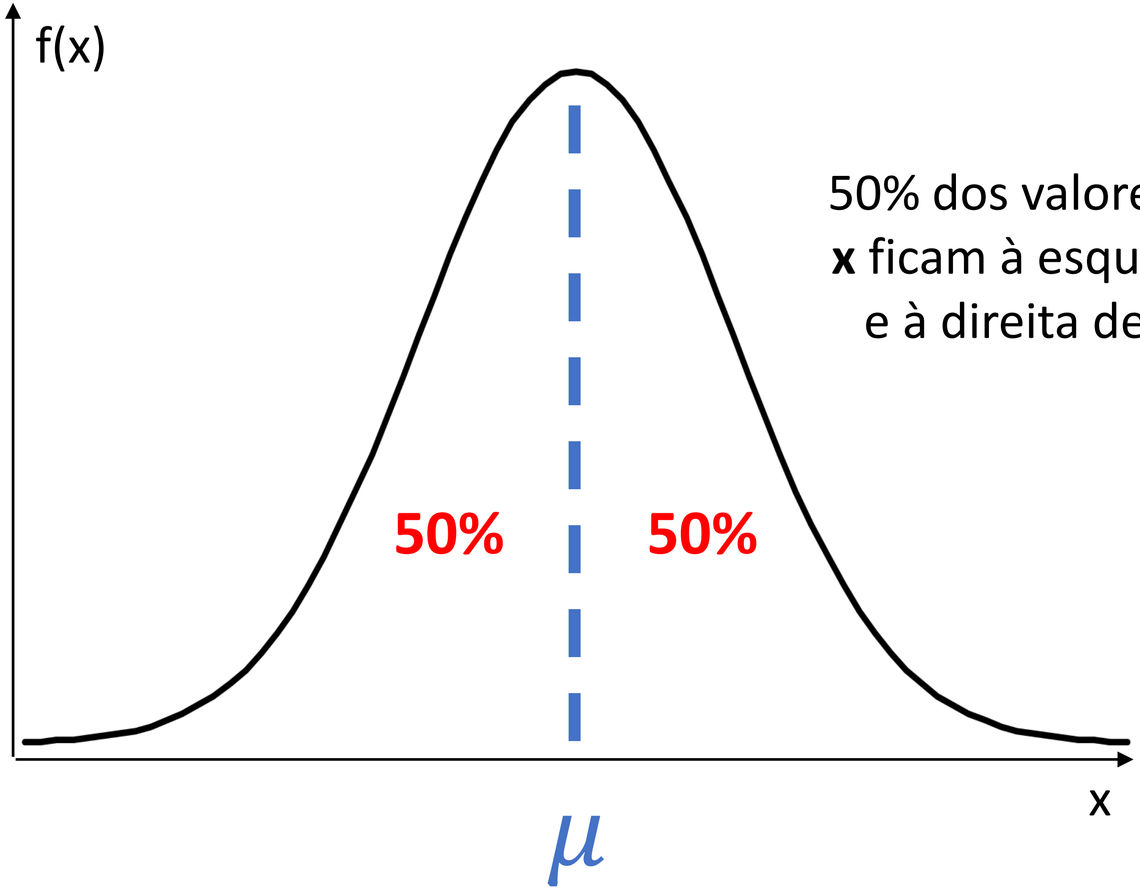
Probabilidades

Variáveis Aleatórias Contínuas: Normal

Algumas propriedades da Função de Densidade **Normal** são facilmente verificadas no seu gráfico. Veja:

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

Propriedade 3: f(x) é simétrica em relação à $\boldsymbol{\mu}$.



50% dos valores de **x** ficam à esquerda e à direita de $\boldsymbol{\mu}$.

Revisão

Vimos que a **Distribuição Normal** pode ser utilizada sempre que o **histograma possuir o formato de sino**.

Ela é a mais famosa e mais utilizada das distribuições de probabilidade porque diversos fenômenos na natureza apresentam uma **Distribuição Normal**, como altura, peso, pressão sanguínea, erros e tantos outros.

Preditiva.ai

Probabilidades

Outras Distribuições de Probabilidades

Probabilidades

Outras Distribuições de Probabilidades

Preditiva.ai

Existem diversas outras **distribuições de probabilidades** utilizadas para calcular a probabilidade dos eventos de interesse em inúmeras situações. Neste quadro resumo mostramos as principais:

Variável Aleatória	Distribuição	Principais aplicações práticas	Link
Discreta	Uniforme	Sorteios de números aleatórios.	https://pt.wikipedia.org/wiki/Distribuição_uniforme
Discreta	Bernoulli e Binomial	Taxa de sucesso em qualquer tipo de processo. Ex.: Vendas, Medicina, Controle de Qualidade etc.	https://pt.wikipedia.org/wiki/Distribuição_binomial
Discreta	Poisson	Contagens diversas: Marketing, Chamadas em um Call Center, veículos em um Pedágio etc.	https://pt.wikipedia.org/wiki/Distribuição_de_Poisson
Discreta	Geométrico	Estudo de cinemática da Física de objetos.	https://pt.wikipedia.org/wiki/Distribuição_geométrica
Discreta	Hipergeométrico	Cálculo da aceitação de lotes em fábricas.	https://pt.wikipedia.org/wiki/Distribuição_hipergeométrica
Contínua	Normal	Pesquisa, Ciências Socias, Física, Medicina, Agricultura, Engenharia, Finanças etc.	https://pt.wikipedia.org/wiki/Distribuição_normal
Contínua	Gama	Contagens em um processo. Ex.: Necessidade dos consumidores.	https://pt.wikipedia.org/wiki/Distribuição_gama
Contínua	Qui Quadrado	Teoria de Qualidade do Ajuste	https://pt.wikipedia.org/wiki/Qui-quadrado
Contínua	Exponencial	Tempo até ocorrer um próximo evento	https://pt.wikipedia.org/wiki/Distribuição_exponencial
Contínua	t-Student	Estimação de média populacional com amostra pequena (menor que 30) e variância desconhecida.	https://pt.wikipedia.org/wiki/Distribuição_t_de_Student
Contínua	F de Fisher	Análise da Variância entre grupos	https://pt.wikipedia.org/wiki/Distribuição_F_de_Fisher-Snedecor

Revisão Geral

Nesta última seção aprendemos o que são **Variáveis Aleatórias**, e seus tipos **discretos** e **contínuos**. Vimos também algumas **distribuições de probabilidade** mais utilizadas em problemas do dia-a-dia, como a **Uniforme**, **Binomial** e **Normal**.

Nas próximas seções, nossa **meta é extrapolar os resultados das análises** que já aprendemos a fazer: estudaremos sobre **Inferência Estatística** e como extrapolar os resultados obtidos a partir de uma **amostra** para a **população**!

Preditiva.ai