DISTRIBUIÇÕES DE PROBABILIDADE

Introdução

Uma *distribuição de probabilidade* é um modelo matemático que relaciona um certo valor da variável em estudo com a sua probabilidade de ocorrência.

Há dois tipos de distribuição de probabilidade:

1. Distribuições <u>Contínuas</u>: Quando a variável que está sendo medida é expressa em uma escala contínua, como no caso de uma característica dimensional.

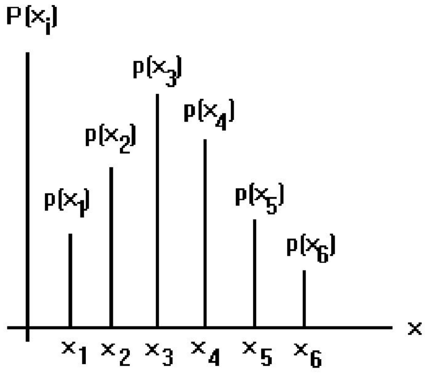
2. Distribuições <u>Discretas</u>: Quando a variável que está sendo medida só pode assumir certos valores, como por exemplo os valores inteiros: 0, 1, 2, etc.

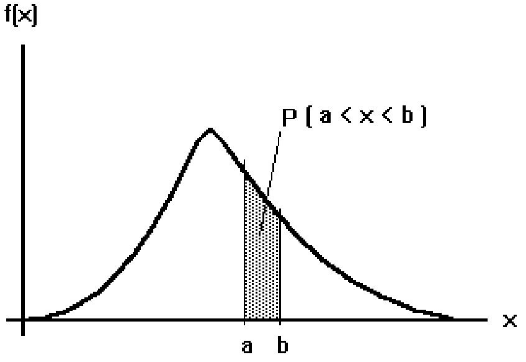
Introdução

No caso de distribuições discretas, a probabilidade de que a variável X assuma um valor específico x_o é dada por: $P(X = x_o) = P(x_o)$

No caso de variáveis contínuas, as probabilidades são especificadas em termos de intervalos, pois a probabilidade associada a um número específico é zero.

$$P(a \leq X \leq b) = \int_a^b f(x)\, dx$$





Distribuições Discretas Mais Importantes

Distribuição de Bernoulli

Distribuição Binomial

Distribuição de Poisson

Distribuição de Bernoulli

Os experimentos mais simples em que observamos a presença ou não de alguma característica são conhecidos como ensaios de Bernoulli. Alguns exemplos:

- **Lançar uma moeda e observar se ocorre cara ou coroa;**
- **Lançar um dado e observar se ocorre seis ou não;**
- **Numa linha de produção, observar se um item, tomado ao acaso, é defeituoso ou não defeituoso;**
- **Verificar se um servidor de intranet está ativo ou não ativo.**

Distribuição de Bernoulli

Denominamos sucesso e fracasso os dois eventos possíveis em cada caso.

O ensaio de Bernoulli é caracterizado por uma variável aleatória X, definida por X=1, se sucesso; X=0, se fracasso.

A função de probabilidade de X (Distribuição de Bernoulli) é dada por

Onde p= P{Sucesso}

x	*p(x)*
0	1-p
1	p
Total	1

Distribuição Binomial

A distribuição binomial é adequada para descrever situações em que os resultados de uma variável aleatória podem ser agrupados em apenas duas classes ou categorias.

As categorias devem ser mutuamente excludentes, de forma que não haja dúvidas na classificação do resultado da variável nas categorias e coletivamente exaustivas, de forma que não seja possível nenhum outro resultado diferente das categorias.

Por exemplo, um produto manufaturado pode ser classificado como perfeito ou defeituoso, a resposta de um questionário pode ser verdadeira ou falsa, as chamadas telefônicas podem ser locais ou interurbanas.

Distribuição Binomial

Mesmo variáveis contínuas podem ser divididas em duas categorias, como por exemplo, a velocidade de um automóvel pode ser classificada como dentro ou fora do limite legal.

Geralmente, denomina-se as duas categorias como *sucesso* ou *falha*. Como as duas categorias são mutuamente excludentes e coletivamente exaustivas:

$$P(sucesso) + P(falha) = 1$$

Conseqüentemente, sabendo-se que, por exemplo, a probabilidade de sucesso é *P(sucesso) = 0,6*, a probabilidade de falha é *P(falha) = 1-0,6 = 0,4*.

Distribuição Binomial

Condições de aplicação:

- são feitas *n* repetições do experimento, onde n é uma constante;
- há apenas dois resultados possíveis em cada repetição, denominados sucesso e falha
- a probabilidade de sucesso *(p)* e de falha *(1- p)* permanecem constante em todas as repetições;
- as repetições são independentes, ou seja, o resultado de uma repetição não é influenciado por outros resultados.

Distribuição Binomial

Seja um processo composto de uma seqüência de *n* observações independentes com probabilidade de sucesso constante igual a *p*, a distribuição do número de sucessos seguirá o modelo Binomial:

$$P(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \mathbf{x = 0,1,....,n}$$

onde $\binom{n}{x}$ **representa o número de combinações de *n* objetos tomados *x* de cada vez, calculado como:**

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

Distribuição Binomial

Os parâmetros da distribuição Binomial são *n* e *p*.

A média e a variância são calculadas como:

$$\mu = np$$

$$\sigma^2 = np(1 - p)$$

A distribuição Binomial é usada com freqüência no controle de qualidade quando a amostragem é feita sobre uma população infinita ou muito grande.

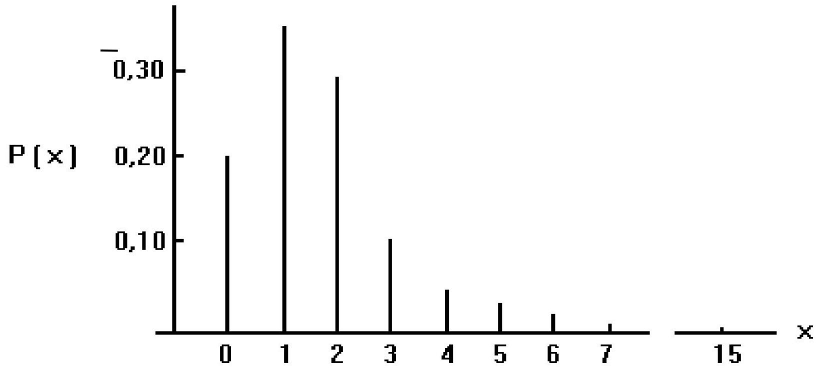
Nas aplicações de controle da qualidade, *x* em geral representa o número de defeituosos observados em uma amostra de *n* itens.

Distribuição Binomial

Por exemplo, se *p = 0,10* e *n = 15*, a probabilidade de obter *x* itens não conformes é calculada usando a equação da Binomial. Por exemplo, para *x=1*

$$\binom{15}{1} = \frac{15!}{1!(15-1)!} = 15$$

$$\hat{P}(1) = \binom{15}{1} x\ 0{,}10^1\ x\ (1-0{,}10)^{15-1} = 15 \times 0{,}10\ x\ 0{,}23 = 0{,}34$$



Distribuição Binomial

Distribuições binomiais com p=0,5 são simétricas. A assimetria aumenta à medida que p aproximasse de zero (assimetria positiva) ou de um (assimetria negativa)

Distribuição de Poisson

A distribuição de Poisson é adequada para descrever situações onde existe uma probabilidade de ocorrência em um campo ou intervalo contínuo, geralmente tempo ou área.

Por exemplo, o nº de acidentes por mês, nº de defeitos por metro quadrado, nº de clientes atendidos por hora.

Nota-se que a variável aleatória é discreta (número de ocorrência), no entanto a unidade de medida é contínua (tempo, área).

Além disso, as falhas não são contáveis, pois não é possível contar o número de acidentes que não ocorreram, nem tampouco o número de defeitos que não ocorreram.

Distribuição de Poisson

Condições de aplicação:

- o número de ocorrências durante qualquer intervalo depende somente da extensão do intervalo;
- as ocorrências ocorrem independentemente, ou seja, um excesso ou falta de ocorrências em algum intervalo não exerce efeito sobre o número de ocorrências em outro intervalo;
- a possibilidade de duas ou mais ocorrências acontecerem em um pequeno intervalo é muito pequena quando comparada à de uma única ocorrência.

Distribuição de Poisson

A distribuição de Poissson fica completamente caracterizada por um único parâmetro λ que representa a taxa média de ocorrência por unidade de medida.

A equação para calcular a probabilidade de x ocorrências é dada por:

$$P(x) = \frac{e^{-\lambda}\lambda^{x}}{x!} \quad x = 0, 1, \ldots, n$$

A média e a variância da distribuição de Poisson são:

$$\mu = \lambda$$

$$\sigma^2 = \lambda^2$$

Distribuição de Poisson

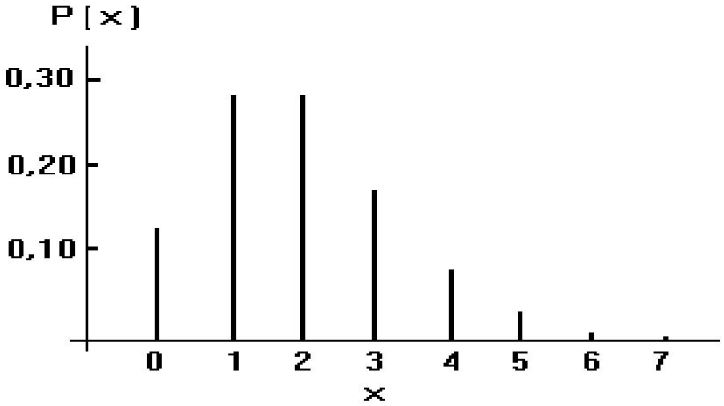
A aplicação típica da distribuição de Poisson no controle da qualidade é como um modelo para o número de defeitos (não-conformidades) que ocorre por unidade de produto (por m^2, por volume ou por tempo, etc.).

Distribuição de Poisson

O número de defeitos de pintura segue uma distribuição de Poisson com $\lambda = 2$.

Então, a probabilidade que uma peça apresente mais de 4 defeitos de pintura virá dada por:

$$1 - P\{X \le 4\} = 1 - \sum^{4} \frac{e^{-2} 2^{4}}{4!} = 1 - 0{,}945 = 0{,}055 = 5{,}5\%$$



x	P(x)
0	0,135
1	0,270
2	0,270
3	0,180
4	0,090
5	0,036
6	0,012

Distribuições contínuas

Distribuição Exponencial

Distribuição Normal

Distribuição Exponencial

Na distribuição de Poisson, a variável aleatória é definida como o número de ocorrências em determinado período, sendo a média das ocorrências no período definida como λ.

Na distribuição Exponencial a variável aleatória é definida como o tempo entre duas ocorrências, sendo a média de tempo entre ocorrências de $1/\lambda$.

Por exemplo, se a média de atendimentos no caixa bancário é de λ = 6/min, então o tempo médio entre atendimentos é $1/\lambda$ = 1/6 de minuto ou 10 segundos.

Distribuição Exponencial

O modelo da distribuição Exponencial é o seguinte:

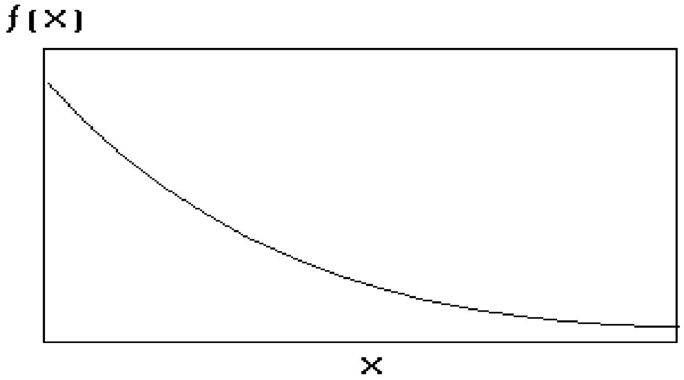
$$f(t) = \lambda e^{-\lambda t}; \qquad t \geq 0$$

onde $\lambda > 0$ é uma constante.

A média e o desvio padrão da distribuição exponencial são calculados usando:

$$\mu = \frac{1}{\lambda}$$

$$\sigma = \frac{1}{\lambda}$$



Distribuição Exponencial

Por exemplo, suponha que uma máquina falhe em média uma vez a cada dois anos λ=*1/2=0,5*. Calcule a probabilidade da máquina falhar durante o próximo ano.

$$F(t) = P\{T \leq 1\} = 1 - e^{0,5x1} = 1 - 0,607 = 0,393$$

A probabilidade de falhar no próximo ano é de *0,393* e de não falhar no próximo ano é de *1-0,393=0,607*.

Ou seja, se forem vendidos 100 máquinas 39,3% irão falhar no período de um ano.

Conhecendo-se os tempos até a falha de um produto é possível definir os períodos de garantia.

Distribuições Normal

A distribuição Normal é a mais importante das distribuições estatísticas, tanto na teoria como na prática:

- **Representa a distribuição de freqüência de muitos fenômenos naturais;**
- **Serve como aproximação da distribuição Binomial, quando *n* é grande;**
- **As médias e as proporções de grandes amostras seguem a distribuição Normal (Teorema do Limite Central).**

Distribuições Normal

A distribuição Normal é em forma de sino, unimodal, simétrica em relação à sua média e tende cada vez mais ao eixo horizontal à medida que se afasta da média.

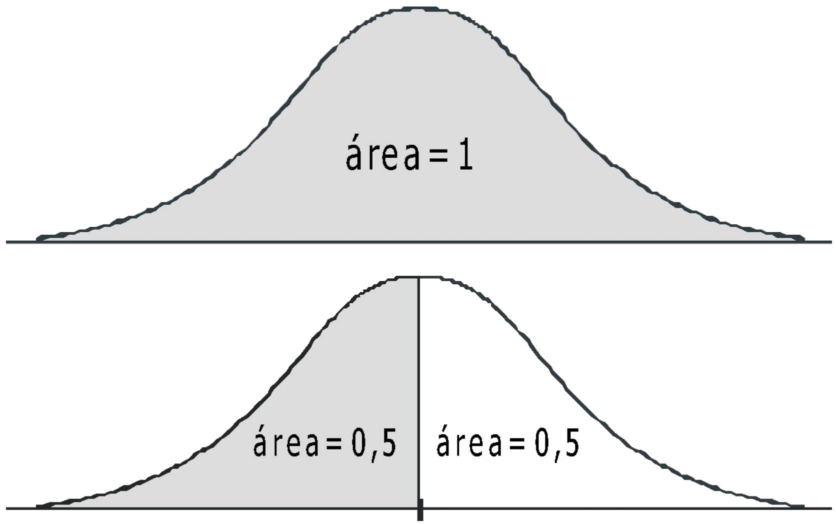
Ou seja, teoricamente os valores da variável aleatória podem variar de -∝ a +∝.

A área abaixo da curva Normal representa 100% de probabilidade associada a uma variável.

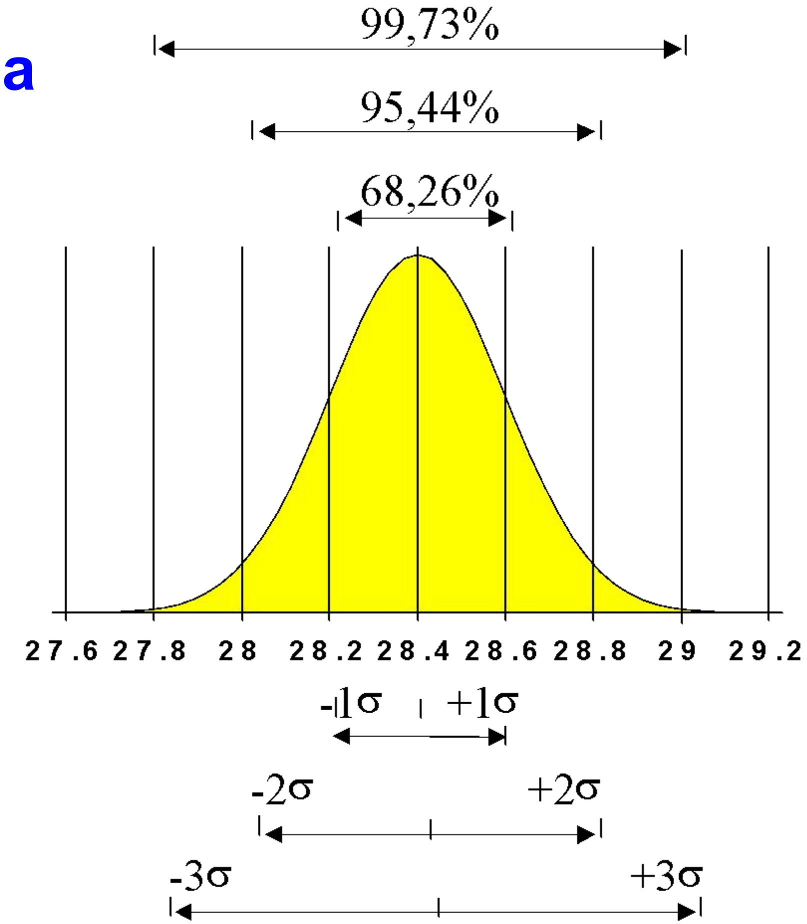
A probabilidade de uma variável aleatória tomar um valor entre dois pontos quaisquer é igual à área compreendida entre esses dois pontos.

Distribuições Normal

A área total abaixo da curva é considerada como 100%.
Isto é, a área total abaixo da curva é 1.



Percentuais da distribuição Normal:



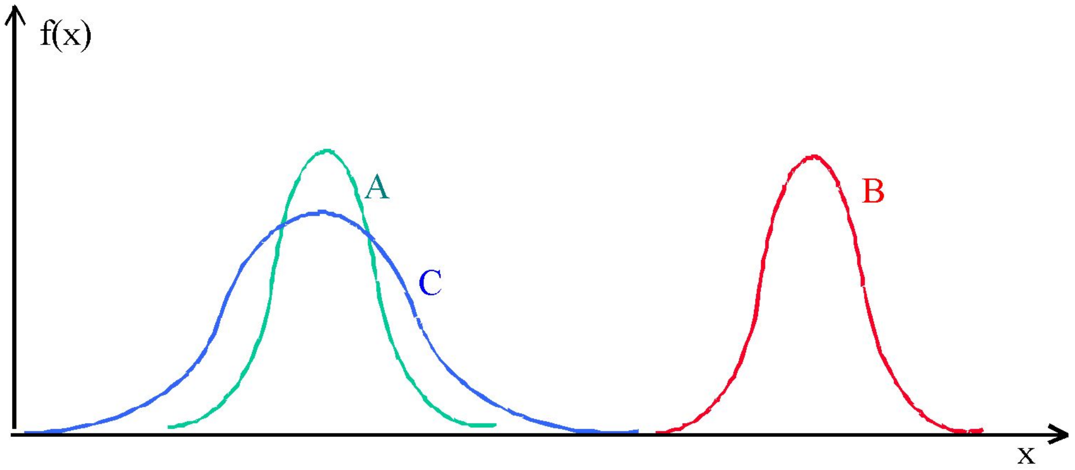
Distribuições Normal

A distribuição Normal fica completamente caracterizada por dois parâmetros: a média e o desvio-padrão.

Ou seja, diferentes médias e desvio-padrões originam curvas normais distintas, como se pode visualizar nos exemplos contidos na tabela abaixo onde há amostras provenientes de distribuições com média e desvios-padrões distintos.

Amostras	Dados	Localização ($\bar{x}$)	Variabilidade ($\bar{R}$)
A	10 12 14 16 18	$\bar{x} = 14$	$\bar{R} = 8$
B	22 24 26 28 30	$\bar{x} = 26$	$\bar{R} = 8$
C	6 10 14 18 22	$\bar{x} = 14$	$\bar{R} = 16$

Distribuições Normal



a) da distribuição A para B muda a tendência central, mas a variabilidade é constante;

b) da distribuição A para C muda a variabilidade, mas a tendência central é constante;

c) da distribuição B para C muda a tendência central e a variabilidade.

Distribuições Normal

Uma conseqüência importante do fato de uma distribuição Normal ser completamente caracterizada por sua média e desvio-padrão é que a área sob a curva entre um ponto qualquer e a média é função somente do número de desvios-padrões que o ponto está distante da média.

Como existem uma infinidade de distribuições normais (uma para cada média e desvio-padrão), transformamos a unidade estudada seja ela qual for (peso, espessura, tempo, etc.) na unidade Z, que indica o número de desvios-padrão a contar da média.

Distribuição Normal

A variável reduzida mede a magnitude do desvio em relação à média, em unidades de desvio padrão.

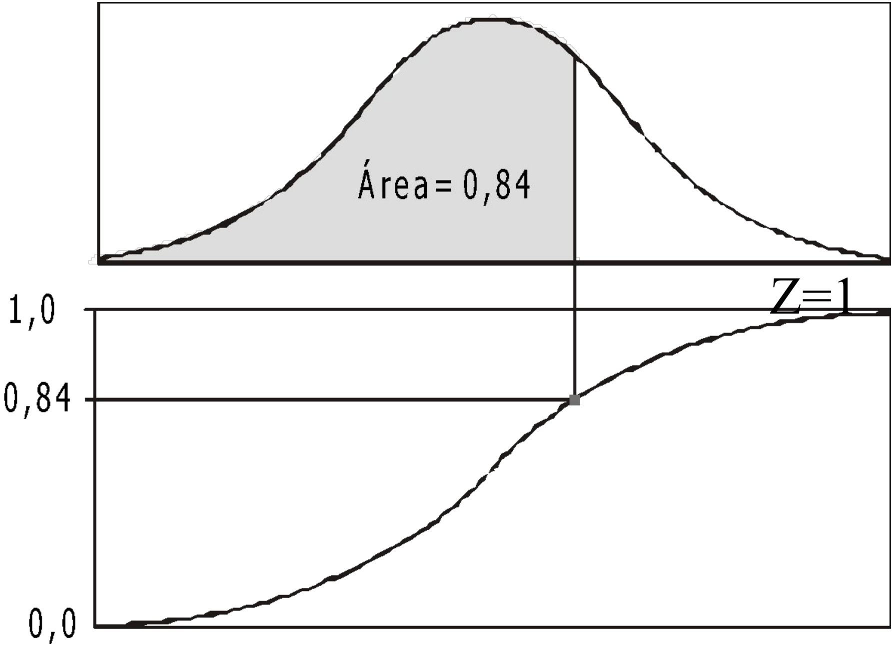
***Z* = *1,5* significa uma observação está desviada 1,5 desvios padrão para cima da média.**

A variável reduzida é muito útil para comparar distribuições e detectar dados atípicos.

Dados são considerados atípicos quando *Z* > *3*.

$$Z = \frac{X - \overline{X}}{S}$$

Para sabermos o valor da probabilidade, utilizamos a tabela da distribuição Normal. Essa tabela nos fornece a área acumulada até o valor de Z



Distribuições Normal

As áreas correspondentes as probabilidades da distribuição normal padrão estão tabeladas.

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.0	**0.8413**	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	**0.8770**	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9278	0.9292	0.9306	0.9319

Probabilidade de ocorrência de valores abaixo de Z

Distribuições Normal

O cálculo da variável reduzida *Z* faz uma transformação dos valores reais em valores codificados.

Essa transformação é feita descontando-se a média para eliminar o efeito de localização (tendência central) e dividindo-se pelo desvio-padrão para eliminar o efeito de escala (variabilidade).

Uma vez calculada a variável reduzida *Z*, consulta-se a tabela Normal padronizada para identificar a probabilidade acumulada à esquerda de *Z*, ou seja, a probabilidade de ocorrerem valores menores ou iguais a um certo valor de *Z* consultado.

Distribuição Normal

Exemplo 1: Suponha que o peso de um rolo de arame seja normalmente distribuído com média 100 e desvio-padrão 10. Então o peso está em torno de 100 a uma distância as vezes maior, as vezes menor que 10.

Queremos saber qual a probabilidade que um rolo, pego ao acaso da produção, possuir peso menor ou igual a 110: P(x <110) = P(Z < 1) = 0,8413

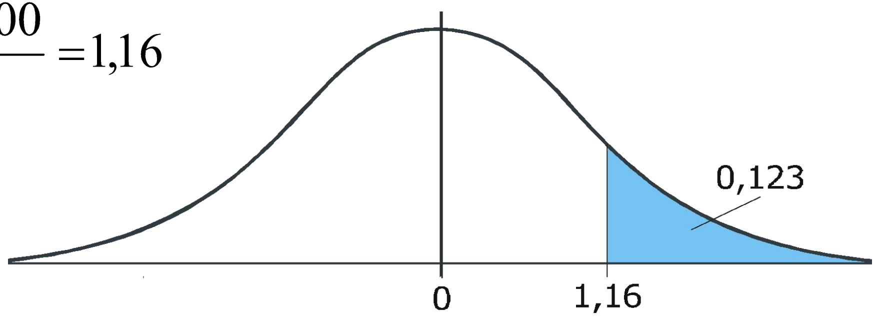
Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.0	**0.8413**	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	**0.8770**	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9278	0.9292	0.9306	0.9319

Probabilidade de ocorrência de valores abaixo de Z

Distribuição Normal

Se quiséssemos saber a probabilidade do peso do rolo ser maior que 111,6, iniciamos calculando o valor de Z:

$$Z = \frac{111{,}6 - 100}{10} = 1{,}16$$



Encontramos o valor de probabilidade 0,8770.

P(Z > 1,16) = 1 - P(Z < 1,16) = 1 - 0,8770 = 0,123

Distribuição Normal

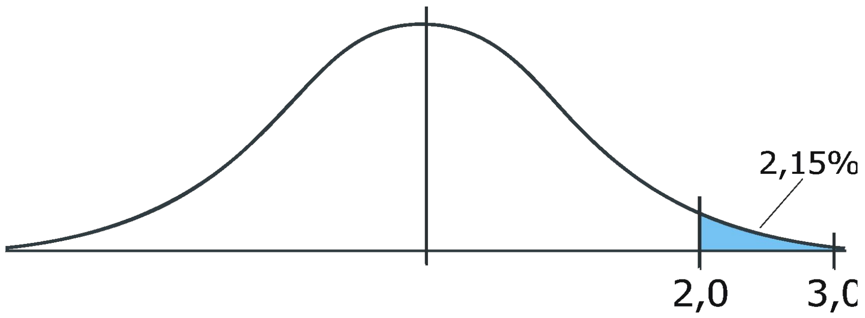
Da mesma forma, se quiséssemos a probabilidade do peso estar entre 120 e 130, teríamos que fazer o seguinte raciocínio:

P(120 < X < 130) = P(X <130) – P(X < 120) =

P(Z< 3) – P(Z< 2) =

0,9987 – 0,9772 = 0,0215

ou seja, 2,15% de chance de um rolo pesar entre 120 e 130



Distribuições Normal

Exemplo 1: **A resistência à tração do papel usado em sacolas de super-mercado é uma característica de qualidade importante.**

Sabe-se que essa resistência segue um modelo Normal com média 40 psi e desvio padrão 2 psi.

Se a especificação estabelece que a resistência deve ser maior que 35 psi, qual a probabilidade que uma sacola produzida com este material satisfaça a especificação?

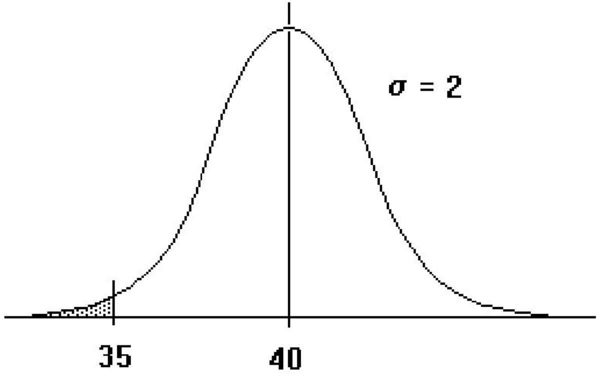
Distribuições Normal

$$P\{X \geq 35\} = 1 - P\{X \leq 35\}$$

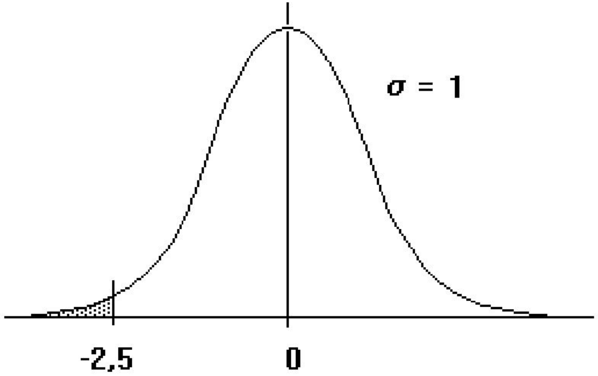
$$P\{X \leq 35\} = P\left\{Z \leq \frac{35-40}{2}\right\} = P\{Z \leq -2{,}5\}$$

Tabela: F(-2,5) = 0,0062

Assim a resposta é 1 - 0,0062 = 99,38%



Distribuição para X (valores reais)



Distribuição para Z (valores codificados)

Distribuições Normal

Exemplo 2: **O diâmetro do eixo principal de um disco rígido segue a distribuição Normal com média 25,08 in e desvio padrão 0,05 in.**

Se as especificações para esse eixo são 25,00 ± 0,15 in, determine o percentual de unidades produzidas em conformidades com as especificações.

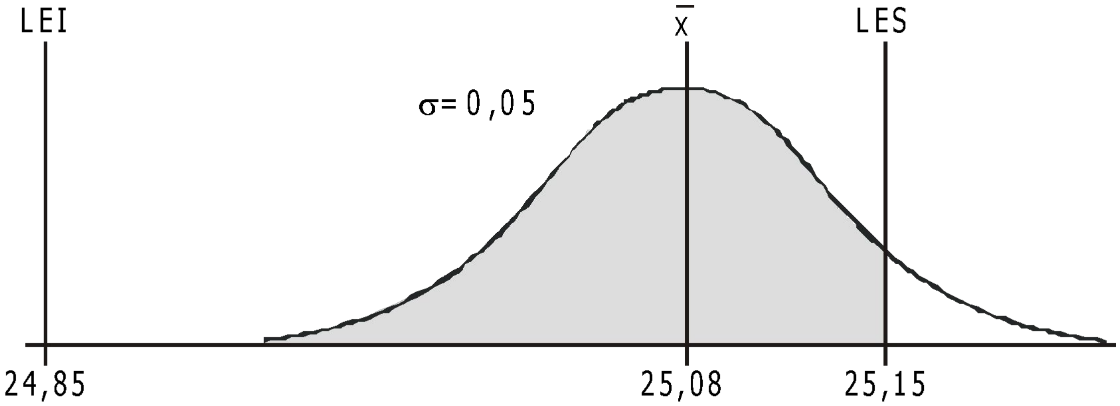
$$P\{24{,}85 \le x \le 25{,}15\} = P\{x \le 25{,}15\} - P\{x \le 24{,}85\}$$

$$= P\left\{Z \le \frac{25{,}15 - 25{,}08}{0{,}05}\right\} - P\left\{Z \le \frac{24{,}85 - 25{,}08}{0{,}05}\right\}$$

$$= P\{Z \le 1{,}40\} - P\{Z \le -4{,}60\} = 0{,}9192 - 0{,}0000 = 0{,}9192$$

Distribuições Normal

ou seja, 91,92% dentro das especificações(área cinza) e 8,08% fora das especificações.



Propriedades da Distribuição Normal

A distribuição Normal tem muitas propriedades úteis.

Uma dessas propriedades é que qualquer combinação linear de variáveis normalmente distribuídas também seguirá o modelo Normal, ou seja:

Se $X_1, X_2, \ldots\ldots\ldots, X_n$ têm distribuição normal e são independentes,

A variável Y que é uma combinação linear de X:

$$Y = a_1X_1 + a_2X_2 + \ldots\ldots + a_kX_k$$

Propriedades da Distribuição Normal

Também seguirá o modelo normal, com média

$$\mu_\gamma = a_1\mu_1 + \ldots\ldots + a_n\mu_n$$

e variância

$$\sigma_Y^2 = a_1^2\sigma_1^2 + \ldots\ldots\ldots + a_n^2\sigma_n^2$$

onde $a_1, \ldots\ldots, a_n$ **são constantes.**

Teorema do Limite Central

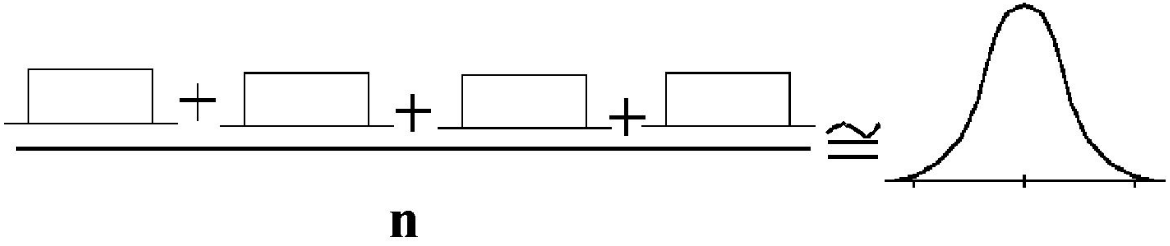
O Teorema do Limite Central indica que a soma (e por conseguinte a média) de *n* variáveis independentes seguirá o modelo Normal, independentemente da distribuição das variáveis individuais.

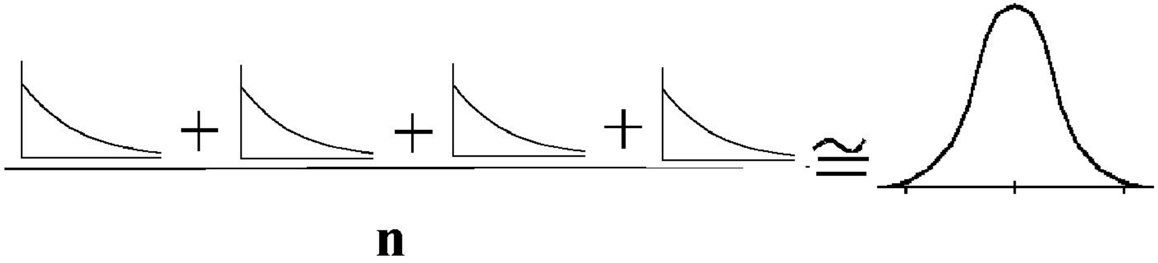
A aproximação melhora na medida em que *n* aumenta. Se as distribuições individuais não são muito diferentes da Normal, basta $n = 4$ ou 5 para se obter uma boa aproximação.

Se as distribuições individuais forem radicalmente diferentes da Normal, então será necessário $n = 20$ ou mais.

Teorema do Limite Central

Na figura abaixo pode ser visto um desenho esquemático do teorema do limite central.





Teorema do Limite Central

Exemplo 5: **A distribuição de probabilidade da variável resultante do lançamento de um dado segue a distribuição uniforme, ou seja, qualquer valor (1,2,3,4,5,6) tem a mesma probabilidade (1/6) de ocorrer.**

No entanto, se ao invés de lançar um dado, sejam lançados dois dados e calculada a média, a média dos dois dados seguirá uma distribuição aproximadamente Normal.

1º dado	2º dado	Soma	Média
1	1	2	1,0
1	2	3	1,5
2	1	3	1,5
1	3	4	2,0
3	1	4	2,0
2	2	4	2,0
1	4	5	2,5
4	1	5	2,5
3	2	5	2,5
2	3	5	2,5
1	5	6	3,0
5	1	6	3,0
2	4	6	3,0
4	2	6	3,0
3	3	6	3,0
1	6	7	3,5
6	1	7	3,5
2	5	7	3,5

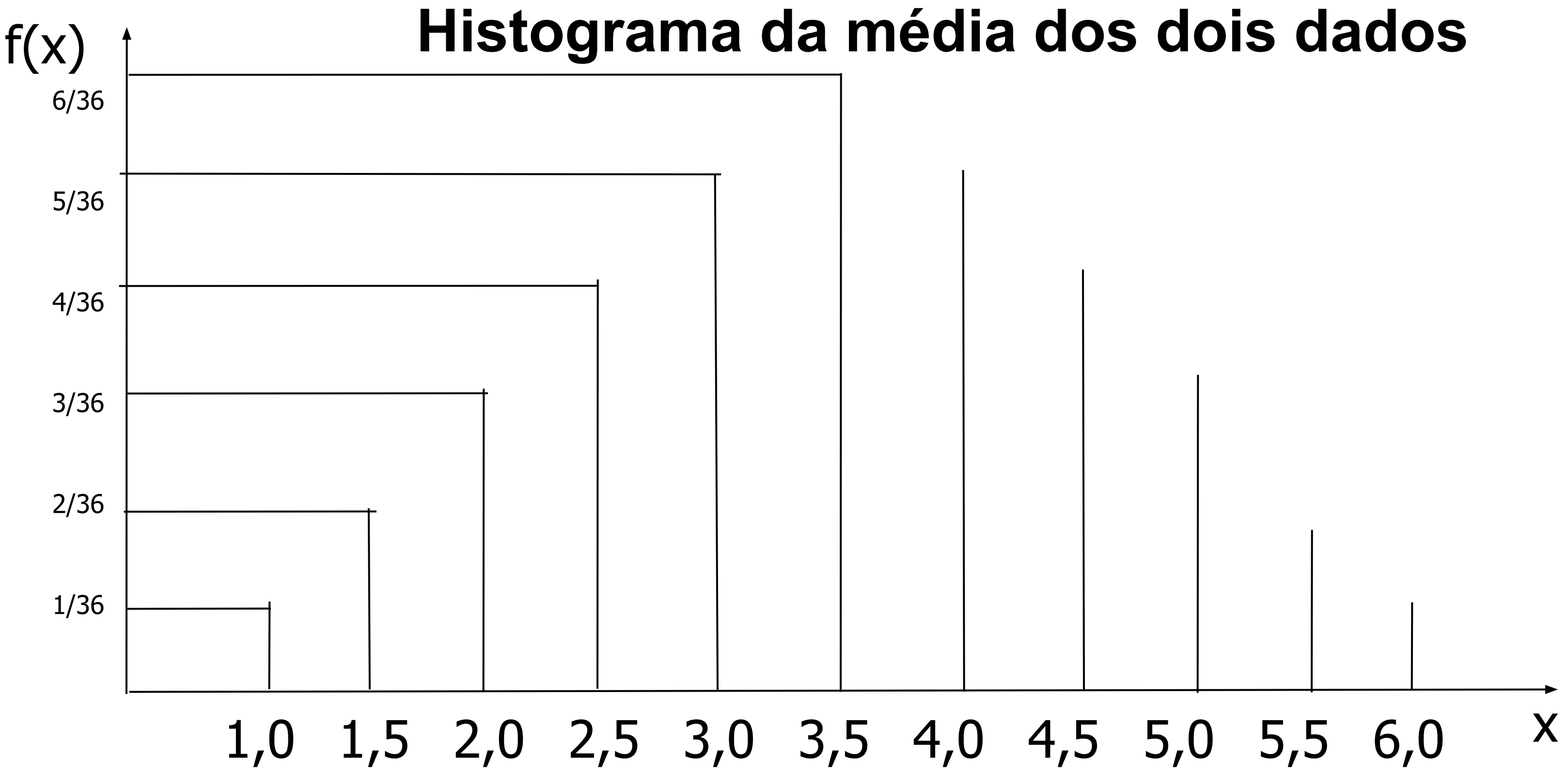
1º dado	2º dado	Soma	Média
5	2	7	3,5
3	4	7	3,5
4	3	7	3,5
2	6	8	4,0
6	2	8	4,0
3	5	8	4,0
5	3	8	4,0
4	4	8	4,0
3	6	9	4,5
6	3	9	4,5
4	5	9	4,5
5	4	9	4,5
4	6	10	5,0
6	4	10	5,0
5	5	10	5,0
5	6	11	5,5
6	5	11	5,5
6	6	12	6,0

Teorema do Limite Central

Tabela de freqüência da média dos dois dados

Média de dois dados	Freqüência
1,0	1
1,5	2
2,0	3
2,5	4
3,0	5
3,5	6
4,0	5
4,5	4
5,0	3
5,5	2
6,0	1

Teorema do Limite Central



Teorema do Limite Central

Conforme pode ser visto no histograma anterior, o histograma da média dos dois dados resulta aproximadamente Normal. Além disso, observa-se que a aproximação da distribuição Normal melhora na medida que se fizesse a média do lançamento de mais dados.