ESTATÍSTICA

Michelle Hanne Soares de Andrade
michellehanne.andrade@gmail.com

Padronizando uma Variável Aleatória

- suponha que, dada uma variável aleatória X, definamos uma nova variável aleatória subtraindo sua média μ e dividindo o resultado por seu desvio-padrão σ.

$$Z \equiv \frac{X-\mu}{\sigma},$$

- que podemos escrever como $Z = aX + b$, onde $a \equiv (1/\sigma)$ e $b \equiv -(\mu/\sigma)$. Então, de acordo com a propriedade 2 do Valor Esperado,

$$\mathrm{E}(Z) = a\mathrm{E}(X) + b = (\mu/\sigma) - (\mu/\sigma) = 0.$$

Padronizando uma Variável Aleatória

- Da propriedade Variância 2,

$$\mathrm{Var}(Z) = a^2\mathrm{Var}(X) = (\sigma^2/\sigma^2) = 1.$$

- Portanto, a variável aleatória *Z* tem uma média zero e uma variância (e portanto um desvio-padrão) igual a um. Esse procedimento algumas vezes é referido como *padronização* da variável aleatória *X*, e *Z* é chamado uma **variável aleatória padronizada.**
- **Frequentemente utilizada na Inferência Estatística.**
- **Exemplo:** Como um exemplo específico, suponha que E(*X*) = 2 e Var(*X*)= 9. Então, *Z*= (*X*-2)/3 terá um valor esperado igual a zero e variância igual a um.

Distribuição de Bernoulli

Distribuição de Bernoulli

Definição

Experimentos que admitem apenas dois resultados possíveis (sucesso ou fracasso) recebem o nome de ensaios de Bernoulli e originam uma variável aleatória com distribuição de Bernoulli.

Função de probabilidade

Seja X uma variável aleatória com distribuição de Bernoulli com probabilidade de sucesso p, em que $X = 1$ se o resultado é sucesso e $X = 0$ se o resultado é fracasso. Então, a função de probabilidade de X fica dada por

$$P(X = x) = p^x (1 - p)^{(1-x)},$$

em que $x = 0, 1$.

Ensaios de Bernoulli

Exemplos

- resultado da inspeção de uma peça, defeituosa ou não defeituosa
- opinião de um eleitor, favorável ou outra opinião
- resultado de um exame vestibular, aprovado ou não aprovado
- intenção de voto de um eleitor, partido A ou outra preferência
- assinatura de TV digital, sim ou não
- conclusão de uma corrida para pedestres, sim ou não
- pressão arterial de um paciente, normal ou alterada
- hábito de práticas esportivas, sim ou não

Distribuição de Bernoulli

Esperança

A esperança (ou valor médio) da distribuição de Bernoulli é dada por

$$\begin{aligned} E(X) &= 0 \times P(X = 0) + 1 \times P(X = 1) \\ &= 0 \times (1 - p) + 1 \times p = p. \end{aligned}$$

Variância

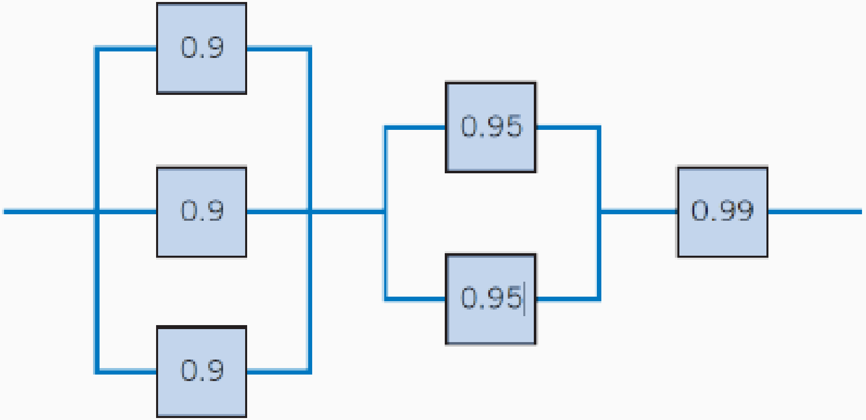
A variância de X é definida por $\mathrm{Var}(X) = E(X^2) - [E(X)]^2$. Temos que

$$\begin{aligned} E(X^2) &= 0^2 \times P(X = 0) + 1^2 \times P(X = 1) \\ &= 0 \times (1 - p) + 1 \times p = p. \end{aligned}$$

Assim, $\mathrm{Var}(X) = p - p^2 = p(1 - p)$ e portanto $\mathrm{DP}(X) = \sqrt{p(1 - p)}$.

Exemplo

- O seguinte circuito opera corretamente se houver um caminho funcional de dispositivos da esquerda para a direita.
- As probabilidades de cada dispositivo funcionar corretamente são dadas na figura abaixo



- Assumindo que os dispositivos funcionam de forma independente, descreva o funcionamento do circuito através de uma variável aleatória.

Exemplo

- O problema pode ser modelado como uma v.a. de Bernoulli:
  - o sistema opera normalmente (sucesso): $X = 1$;
  - o sistema apresenta falhas (fracasso): $X = 0$.
- Neste caso,

$$P(X = 1) = (1 - 0,1^3)(1 - 0,05^2)(0,99) = 0,9870$$

  e

$$P(X = 0) = 1 - P(X = 1) = 0,0130.$$

- Portanto, temos uma v.a. de Bernoulli com função de probabilidade

$$f(x_i) = 0,9870^{x_i} \cdot 0,0130^{1-x_i}, \quad x_i = 0,1.$$

Distribuição Binomial

Distribuição Binomial

Motivação

Um dado é lançado 3 vezes de forma independente. Qual a probabilidade de obter a face 5 duas vezes?

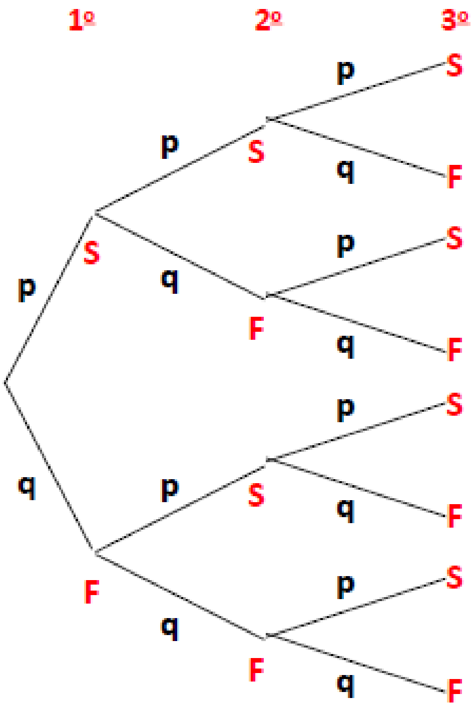
Denotando S como sendo sucesso (obter face 5 num lançamento) e F como sendo fracasso, o espaço amostral pode ser representado por

$$\Omega = \{(SSS), (SSF), (SFS), (FSS), (SFF), (FSF), (FFS), (FFF)\}.$$

Vamos considerar a variável aleatória X: número de sucessos nos três lançamentos, sendo $p = P(\mathrm{S})$ e $q = 1 - p = P(\mathrm{F})$ em cada lançamento.

Distribuição Binomial

Diagrama de Árvore



Ω	Prob	X
(SSS)	p^3	3
(SSF)	p^2q	2
(SFS)	p^2q	2
(SFF)	pq^2	1
(FSS)	p^2q	2
(FSF)	pq^2	1
(FFS)	pq^2	1
(FFF)	q^3	0

Distribuição Binomial

Distribuição de probabilidade

Portanto, a função de probabilidade da variável aleatória X: número de sucessos nos três lançamentos fica dada por

x	0	1	2	3
$P(X = x)$	q^3	$3pq^2$	$3p^2q$	p^3

Assim, a função de probabilidade de X pode ser expressa na forma

$$P(X = x) = \binom{3}{x} p^x (1 - p)^{(3-x)},$$

para $x = 0, 1, 2, 3$.

Distribuição Binomial

Distribuição de probabilidade

Em particular, para um dado equilibrado $p = \frac{1}{6}$ obtemos

x	0	1	2	3
$P(X = x)$	$0,5787$	$0,3472$	$0,0694$	$0,0046$

Assim, a probabilidade da face 5 aparecer duas vezes (para um dado equilibrado) fica dada por $P(X = 2) = 0,0694$.

Distribuição Binomial

- A **variável aleatória X** correspondente ao número de sucessos em n ensaios de Bernoulli independentes (no sentido probabilístico) e com mesma **probabilidade p** de sucesso em cada ensaio, tem distribuição binomial **com parâmetros n e p.**
- A função de probabilidade de X é expressa na forma:

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

- em que x = 0, 1,.... n. Denotamos $X \sim B(n, p)$.

Distribuição Binomial

Esperança

Se $X \sim \mathrm{B}(n, p)$ podemos escrever $X = X_1 + \cdots + X_n$, em que $X_i \sim \mathrm{Be}(p)$ para $i = 1, \ldots, n$. Assim, obtemos

$$\mu = \mathrm{E}(X) = \mathrm{E}(X_1) + \cdots + \mathrm{E}(X_n) = np.$$

Variância

Similarmente como temos n ensaios independentes, então

$$\sigma^2 = \mathrm{Var}(X) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n) = np(1-p).$$

E daí segue que $\sigma = \mathrm{DP}(X) = \sqrt{np(1-p)}$.

Distribuição Binomial - Exemplo

Aplicação

Considere uma prova com 12 questões, cada uma com 4 alternativas. Suponha que o aluno escolha a resposta ao acaso. Qual é a probabilidade de que ele acerte pelo menos 6 questões?

Vamos considerar a variável aleatória X: número de questões que o aluno acerta. Vamos supor que $X \sim \mathrm{B}(n, p)$, em que $n = 12$ e $p = 0,25$.

Distribuição Binomial - Exemplo

Aplicação

Portanto, a função de probabilidade de X fica dada por

$$P(X = x) = \binom{12}{x} 0,25^x 0,75^{(12-x)},$$

em que $x = 0, 1, \ldots, 12$. Portanto, usando uma tabela binomial obtemos $P(X \geq 6) = P(X = 6) + \cdots + P(X = 12) \cong 0,0544$.

Adicionalmente, temos que o valor esperado de X fica dado por $\mu = n \times p = 12 \times 0,25 = 3$. Ou seja, espera-se que o aluno acerte 3 questões.

Distribuição Geométrica

Distribuição Geométrica

Definição

Supor que X representa o número de ensaios independentes até a ocorrência do primeiro sucesso que ocorre com probabilidade p. A função de probabilidade de X fica dada por

$$P(X = x) = p(1 - p)^{(x-1)},$$

em que $x = 1, 2, \ldots$. Denotamos $X \sim \mathrm{G}(p)$. É um exemplo de variável aleatória discreta com um número enumerável de valores.

- Valor esperado
  $\mu = \mathrm{E}(X) = \frac{1}{p}$
- Variância
  $\sigma^2 = \mathrm{Var}(X) = \frac{1-p}{p^2}$, logo $\mathrm{DP}(X) = \frac{\sqrt{1-p}}{p}$

Distribuição Geométrica - Exemplo

Aplicação

Num jogo a probabilidade de um jogador ganhar algum prêmio em cada tentativa é de $0,10$. Supondo tentativas independentes, qual a probabilidade do jogador ganhar algum prêmio antes de 5 tentativas? Seja X: número de tentativas até a ocorrência do primeiro sucesso (ganhar algum prêmio). Vamos supor que $X \sim G(0,10)$.

Distribuição Geométrica - Exemplo

Aplicação

Portanto, queremos saber $P(X \leq 4) = \sum_{x=1}^{4} P(X = x)$, em que

$$P(X = x) = p(1-p)^{(x-1)} = 0,10 \times 0,90^{(x-1)}.$$

para $x = 1, 2, 3, 4$.

Daí obtemos

$$\begin{aligned} P(X \leq 4) &= 0,10 \times \{0,90^0 + 0,90^1 + 0,90^2 + 0,90^3\} \\ &= 0,10 \times \{1 + 0,9 + 0,81 + 0,729\} \\ &= 0,10 \times 3,439 \\ &\cong 0,344(34,4\%). \end{aligned}$$

Distribuição de Poisson

- O modelo de Poisson é amplamente utilizado para a contagem de elementos ao longo de um domínio contínuo: número de ligações por minuto, número de casos de dengue por unidade de área, etc.
- Surge quando o número de ensaios em um modelo binomial tende ao infinito (sendo a probabilidade de sucesso mantida constante).
- *X* é uma v.a. aleatória que pode assumir os valores *k* = 0, 1, 2, 3,...
- Função de probabilidade:

$$
\begin{aligned}
f(k) &= P(X = k) \\
&= \lim_{n \to +\infty} \binom{n}{k} p^k (1-p)^{n-k}.
\end{aligned}
$$

Distribuição de Poisson

- A função de probabilidade pode ser simplificado se trocarmos o conceito de probabilidade de sucesso pela ideia de taxa média de sucesso.
- Seja $\lambda = np$ o número médio de sucessos ao longo de n ensaios de Bernoulli. Neste caso, $p = \lambda/n$ e:

$$
\begin{aligned}
f(k) &= \lim_{n \to +\infty} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
&= \lim_{n \to +\infty} \frac{n(n-1)\cdots(n-k+1)}{n \cdot n \cdots n \cdot k!} \cdot \lambda^k \cdot \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \\
&= 1 \cdot \frac{1}{k!} \cdot \lambda^k \cdot e^{-\lambda} \cdot 1 = \frac{e^{-\lambda} \cdot \lambda^k}{k!}.
\end{aligned}
$$

- Desta forma, o modelo de Poisson é completamente especificado pelo parâmetro λ (a taxa média de sucessos esperados).

Distribuição de Poisson

Definição

Se X representa o número de ocorrências de um evento no tempo ou no espaço e se X segue distribuição de Poisson de parâmetro λ, então a função de probabilidade de X fica dada por

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

em que $x = 0, 1, \ldots$. Denotamos $X \sim \mathrm{P}(\lambda)$. Temos também aqui uma variável aleatória discreta com um número enumerável de valores.

- Valor esperado
  $\mu = \mathrm{E}(X) = \lambda$
- Variância
  $\sigma^2 = \mathrm{Var}(X) = \lambda$, logo $\mathrm{DP}(X) = \sqrt{\lambda}$

Distribuição de Poisson

Exemplos

- nº de acidentes numa rodovia num determinado período
- nº de chamadas telefônicas por minuto
- nº de mensagens que chegam a um servidor por minuto
- nº de pedidos de empréstimno num banco num mês
- nº de defeitos num tecido por metro quadrado
- nº de bactérias numa lâmina de microscópio
- nº de automóveis vendidos numa concessionária num dia

Distribuição de Poisson - Exemplo

Aplicação

Sabe-se que em média ocorrem 1,5 acidentes por dia numa rodovia. Qual a probabilidade de ocorrerem dois ou mais acidentes num dia qualquer?.Seja X: número de acidentes num dia na rodovia. Vamos supor que $X \sim \mathrm{P}(1,5)$.

Temos que $P(X \geq 2) = 1 - P(X = 0) - P(X = 1)$,
em que $P(X = 0) = e^{-1,5} = 0,223$ e $P(X = 1) = e^{-1,5} \times 1,5 = 0,335$.
Daí obtemos

$$P(X \geq 2) = 1 - 0,223 - 0,335 = 0,442.$$

Distribuição de Poisson - Exemplo

Aproximação da Binomial para a Poisson

Se $X \sim \mathrm{B}(n, p)$ então para n grande e p pequeno temos que

$$P(X = x) \cong \frac{e^{-\lambda}\lambda^{x}}{x!},$$

em que $\lambda = np$, $x = 0, 1, \ldots, n$ e $np < 10$.

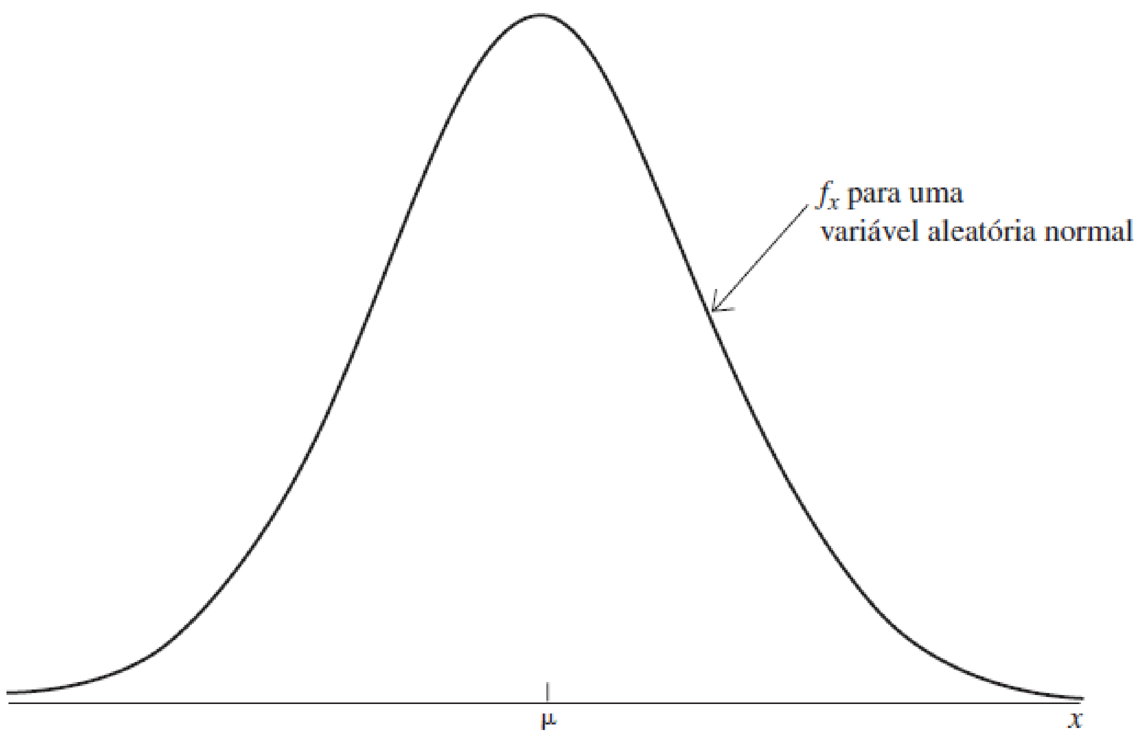
Distribuição Normal

Distribuição Normal

- A distribuição normal e as suas derivadas são as distribuições mais amplamente usadas em estatística. **Assumir que variáveis aleatórias definidas para populações são normalmente distribuídas simplifica os cálculos de probabilidade.**
- Uma variável aleatória **normal é uma variável aleatória contínua** que pode assumir qualquer valor.
- Sua função de densidade de probabilidade tem a forma familiar de um sino traçada conforme imagem.

Distribuição Normal

A forma geral de uma função de densidade de probabilidade normal.



Distribuição Normal

- A distribuição normal e as suas derivadas são as distribuições mais amplamente usadas em estatística. **Assumir que variáveis aleatórias definidas para populações são normalmente distribuídas simplifica os cálculos de probabilidade.**
- Uma variável aleatória **normal é uma variável aleatória contínua** que pode assumir qualquer valor.
- Sua função de densidade de probabilidade tem a forma familiar de um sino traçada conforme imagem.

Distribuição Normal

- Matematicamente, a fdp de *X* pode ser escrita como

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp[-(x-\mu)^2/2\sigma^2], \quad -\infty < x < \infty,$$

- onde $\mu = E(X)$ e $\sigma^2 = Var(X)$. Dizemos que *X* tem uma **distribuição normal** com valor esperado μ e variância σ^2.
- Como a distribuição normal é simétrica em relação a μ, μ também é a mediana de *X*. A distribuição normal é algumas vezes chamada de ***distribuição de Gauss*** em homenagem ao famoso estatístico C. F. Gauss.

Distribuição Normal

Algumas variáveis aleatórias parecem seguir em linhas gerais uma distribuição normal.

Exemplo: As alturas e pesos do ser humano, resultados de provas e índices de desemprego municipais possuem fdps com aproximadamente a forma na anterior.

Outras distribuições, **como as da renda**, não parecem seguir a função de probabilidade normal. Na maioria dos países, a renda **não é simetricamente distribuída em torno de qualquer valor**; a distribuição é inclinada em direção à extremidade superior.

Distribuição Normal

Em alguns casos, uma variável pode ser transformada para atingir a normalidade. **Uma transformação popular é o log natural**, que faz sentido para variáveis aleatórias positivas.

Se X for uma variável aleatória positiva, tal como a renda, e $Y = \log(X)$ tiver uma distribuição normal, então, dizemos que X tem uma ***distribuição lognormal*.**

Exemplo de lognormal: distribuição de renda de muitos países e preços de mercadorias.

A Distribuição Normal Padrão

Um caso especial da distribuição normal ocorre quando a média for zero e a variância (e, portanto, o desvio-padrão) for a unidade. Se uma variável aleatória **Z tiver uma distribuição Normal(0,1)**, dizemos que ela tem um **distribuição normal padrão**.

A Distribuição Normal Padrão

- A fdp de uma variável aleatória normal padrão é representada por PHI(z); com $\mu = 0$ e $\sigma^2 = 1$, ela é dada por:

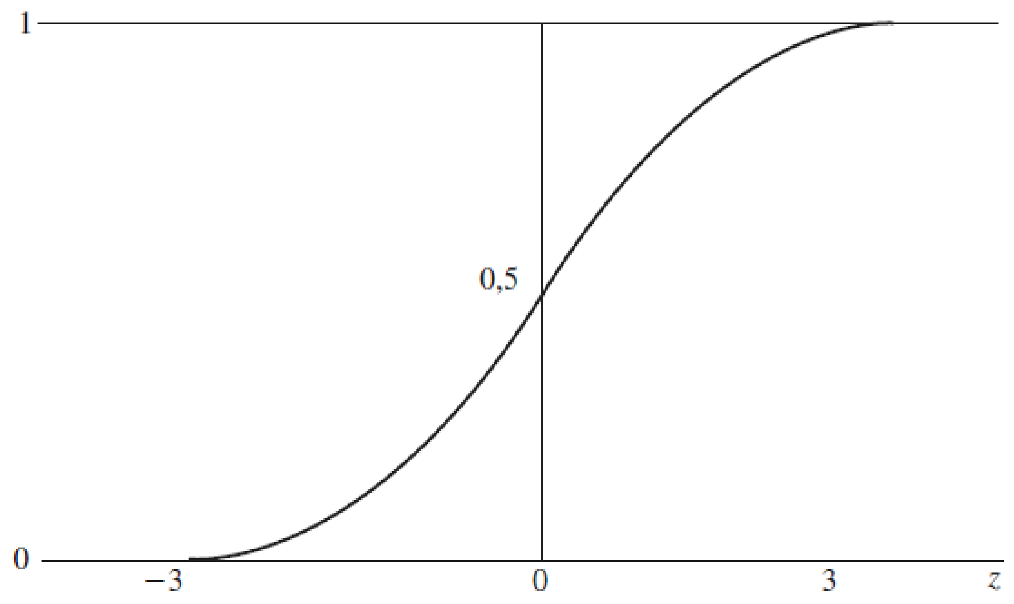
$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp(-z^2/2), \quad -\infty < z < \infty.$$

$\Phi(z)$ é a integral da função.

- **A função de distribuição cumulativa normal** padrão é representada por $\Phi(z)$ e é obtida como a área sob ϕ, à esquerda de z;
- Lembre-se de que $\Phi(z)$= $P(Z \leq z)$; como Z é contínua, $\Phi(z)$= $P(Z < z)$ também é contínua

A Distribuição Normal Padrão

A função de distribuição cumulativa normal padrão.



A Distribuição Normal Padrão

- Os valores de $\Phi(z)$ são facilmente tabulados; eles estão dados para z entre 3,1 e 3,1. Para $z \leq 3{,}1$, $\Phi(z)$ será menor que 0,001, e para $z \geq 3{,}1, \Phi(z)$ será maior que 0,999. Para evitar tabelas grandes, programas de computador fazem esses cálculos.

- ***Usando fatos básicos da probabilidade, podemos usar a fdc normal padrão para calcular a probabilidade de qualquer evento envolvendo uma variável aleatória normal padrão.***

A Distribuição Normal Padrão

$$P(Z > z) = 1 - \Phi(z),$$

$$P(Z < -z) = P(Z > z)$$

E,

$$P(a \leq Z \leq b) = \Phi(b) - \Phi(a).$$