

Distribuições de Probabilidade

Reinaldo Madarazo - 2014

Distribuição de Probabilidades

- Uma distribuição de probabilidades para uma variável aleatória discreta é uma lista mutuamente exclusiva de todos os resultados numéricos possíveis para aquela variável, de modo que uma determinada probabilidade de ocorrência esteja associada a cada resultado.
- **Valor Esperado:** a média aritmética μ de uma distribuição de probabilidades é o valor esperado de sua variável aleatória:

$$\mu = E(X) = \sum_{i=1}^N X_i P(X_i)$$

$X_i = i$ - *ésimo resultado para a variável discreta X*

$P(X_i)$ = *probabilidade de ocorrência do resultado X_i*

- **Variância e Desvio Padrão:**

$$\sigma^2 = \sum_{i=1}^N (X_i - E(X))^2 P(X_i)$$

$X_i = i$ - *ésimo resultado para a variável discreta X*

$P(X_i)$ = *probabilidade de ocorrência do resultado X_i*

$$\sigma = \sqrt{\sigma^2}$$

Exemplo:

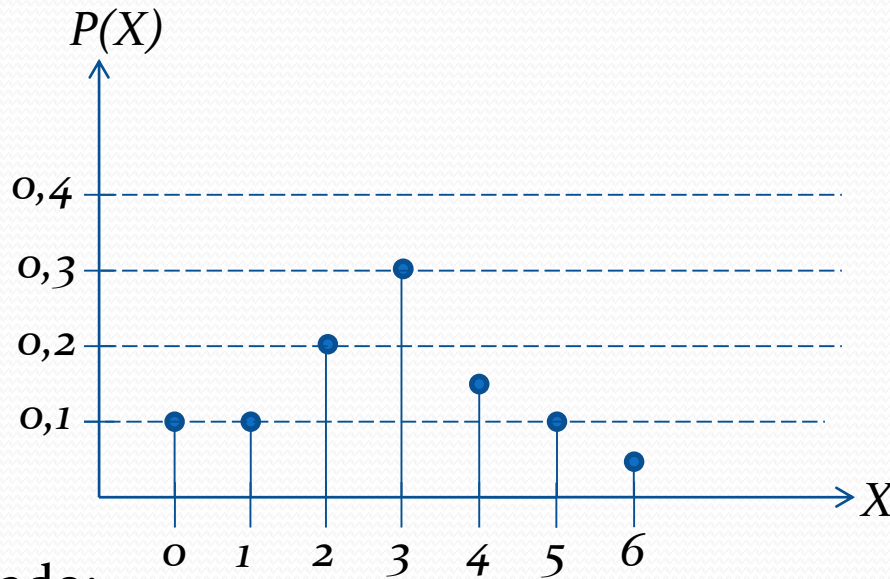
- Considere a tabela a seguir, que mostra o número de *hipotecas aprovadas por semana* e suas *probabilidades de ocorrência*:

Hipotecas Aprovadas	Probabilidade
0	0,10
1	0,10
2	0,20
3	0,30
4	0,15
5	0,10
6	0,05

- 1. Faça o gráfico da Distribuição de probabilidades para as quantidades de hipotecas imobiliárias aprovadas por semana.
- 2. Calcule o valor esperado da quantidade de hipotecas imobiliárias aprovadas por semana.
- 3. Determine a variância e o desvio-padrão da quantidade de hipotecas aprovadas por semana.

Solução

- 1. Gráfico:



- Valor Esperado:

$$\begin{aligned}\mu = E(X) &= \sum_{i=1}^N X_i P(X_i) = (0)(0,1) + (1)(0,1) + (2)(0,2) + \\ &+ (3)(0,3) + (4)(0,15) + (5)(0,1) + (6)(0,05) = 2,8\end{aligned}$$

Solução

- Variância e Desvio-Padrão

<i>Hipotecas (X_i)</i>	<i>$P(X_i)$</i>	<i>$X_i P(X_i)$</i>	<i>$(X_i - E(X_i))^2 P(X_i)$</i>
0	0,10	$(0)(0,10)=0,0$	$(0-2,8)^2(0,10)=0,784$
1	0,10	$(1)(0,10)=0,10$	$(1-2,8)^2(0,10)=0,324$
2	0,20	$(2)(0,20)=0,40$	$(2-2,8)^2(0,20)=0,128$
3	0,30	$(3)(0,30)=0,90$	$(3-2,8)^2(0,30)=0,012$
4	0,15	$(4)(0,15)=0,6$	$(4-2,8)^2(0,15)=0,216$
5	0,10	$(5)(0,10)=0,5$	$(5-2,8)^2(0,10)=0,484$
6	0,05	$(6)(0,05)=0,3$	$(6-2,8)^2(0,05)=0,512$
Soma	1,00	$\mu=E(X)=2,8$	$\sigma^2=2,46 \rightarrow \sigma=1,57$

Distribuição Binomial

- **Modelo Matemático:** é uma expressão matemática que representa uma variável de interesse.
- A **distribuição binomial** é um modelo matemático utilizado quando a variável discreta de interesse é o número de sucessos em uma amostra de n observações e que possui quatro propriedades essenciais:
 - A amostra consiste de um número fixo de observações, n .
 - Cada observação é classificada como uma de duas categorias mutuamente excludentes e coletivamente exaustivas, geralmente chamadas de **sucesso** e **insucesso**.
 - A probabilidade de uma observação ser classificada como sucesso, p , é *constante*, de observação para observação. Por conseguinte, a probabilidade de uma observação ser classificada como insucesso, $1-p$, é também constante de observação para observação.
 - O resultado de qualquer observação é independente do resultado de qualquer outra observação. Para assegurar isso, as observações podem ser selecionadas aleatoriamente, seja a partir de uma *população infinita sem reposição*, seja a partir de uma *população finita com reposição*.

Exemplo

- Quando clientes da *Saxon Home Improvement Company* fazem pedidos pela internet, o sistema de informações contábeis da empresa faz uma revisão nos formulários de pedidos em busca de *possíveis equívocos*.
- Qualquer fatura passível de questionamento é *etiquetada* e incluída em um *relatório de exceções* diário.
- Dados recentes coletados pela empresa mostram que a probabilidade de que um formulário de pedido venha a ser etiquetado é de **0,10**.
- Estamos interessados na *quantidade* de formulários *etiquetados* em uma determinada amostra de pedidos. Suponha que se obtenha o seguinte resultado em uma amostra de 4 pedidos:

Pedido 1	Pedido 2	Pedido 3	Pedido 4
Etiquetado	Etiquetado	Não-Etiquetado	Etiquetado

Exemplo

- Uma vez que a probabilidade histórica de um formulário ser etiquetado é de $p=0,10$, a probabilidade de que cada encomenda venha a ocorrer na sequência é:

Pedido 1	Pedido 2	Pedido 3	Pedido 4
$p=0,10$	$p=0,10$	$1-p=0,90$	$p=0,10$

- Como cada resultado é independente do outro, e como os formulários foram selecionados de uma população extremamente grande (infinita), podemos obter a probabilidade de se ter essa sequência específica:

$$P=p.p.(1-p).p=p^3(1-p)=(0,10)(0,10)(0,10)(0,90)=(0,10)^3(0,90)=0,0009$$

Exemplo

- Esse resultado indica somente a probabilidade de *três pedidos etiquetados (sucesso)*, de uma *amostra de 4 pedidos*, em uma sequência específica.
- Para selecionar X objetos a partir de n objetos, *independente da sequência*, utilizamos **Combinações**:

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

- Com $n=4$ e $X=3$ (3 formulários etiquetados – sucesso) temos:

$$C_{n,X} = \binom{n}{X} = \frac{n!}{X!(n-X)!} = \frac{4!}{3!(4-3)!} = \frac{4.3!}{3!!} = 4$$

Exemplo

- As 4 possibilidades de sequência são:
 - Etiketada, etiketada, etiketada, não-etiketada: $P=p.p.p.(1-p)=p^3(1-p)^1=0,0009$
 - Etiketada, etiketada, não-etiketada, etiketada: $P=p.p.(1-p).p=p^3(1-p)^1=0,0009$
 - Etiketada, não-etiketada, etiketada, etiketada: $P=p.(1-p).p.p=p^3(1-p)^1=0,0009$
 - Não-etiketada, etiketada, etiketada, etiketada: $P=(1-p).p.p.p=p^3(1-p)^1=0,0009$
- A probabilidade total de se obterem 3 formulários etiketados em uma amostra de 4 formulários é *(número de sequências possíveis) x (Probabilidade de uma determinada sequência)* $= (4) \times (0,0009) = 0,0036$
- Assim, a probabilidade de obtermos **3 formulários etiketados** em uma amostra de 4 formulários com probabilidade de sucesso **$p=0,10$** é de **0,0036**.

Modelo Matemático da Distribuição Binomial

- A medida que o número de observações de sucesso vai ficando maior, indo de nenhuma (o) até n , os cálculos envolvidos começam a demandar mais tempo.
- Um modelo matemático proporciona uma fórmula geral para calcular qualquer probabilidade binomial com *probabilidade de sucesso p* e *tamanho da amostra n* e com *número de sucessos X* :

$$P(X) = \frac{n!}{X!(n-X)!} p^X (1-p)^{n-X}$$

$$\text{com } 0 \leq X \leq n$$

Exercícios

- 1) Qual a probabilidade de se obter **três** formulários etiquetados em uma amostra de quatro pedidos?
- 2) Qual a probabilidade de que existam **três ou mais** formulários etiquetados na amostra de quatro pedidos?
- 3) Qual é a probabilidade de que existam **menos de três** formulários de pedidos etiquetados, na amostra de quatro formulários de pedidos?
- Considere para todos os casos a probabilidade de sucesso de **$p=0,10$** .

$$1) P(X = 3) = P(3) = \frac{4!}{3!(4-3)!} (0,10)^3 (1-0,10)^{4-3} = 0,0036$$

2) Para pelo menos 3 pedidos etiquetados temos: $P(X \geq 3) = P(3) + P(4)$

$$P(X = 4) = P(4) = \frac{4!}{4!(4-4)!} (0,10)^4 (1-0,10)^{4-4} = 0,0001$$

$$\text{Assim : } P(X \geq 3) = 0,0036 + 0,0001 = 0,0037$$

Exercícios

3) Para que existam menos de 3 formulários etiquetados : $P(X < 3) = P(0) + P(1) + P(2)$

$$P(X = 0) = P(0) = \frac{4!}{0!(4-0)!} (0,10)^0 (1-0,10)^{4-0} = 0,6561$$

$$P(X = 1) = P(1) = \frac{4!}{1!(4-1)!} (0,10)^1 (1-0,10)^{4-1} = 0,2916$$

$$P(X = 2) = P(2) = \frac{4!}{2!(4-2)!} (0,10)^2 (1-0,10)^{4-2} = 0,0486$$

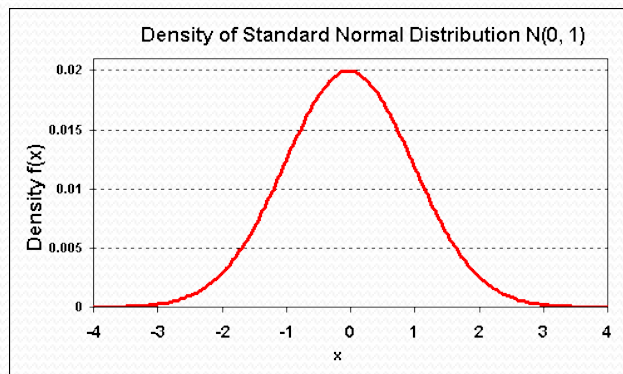
$$\text{Assim : } P(X < 3) = 0,6561 + 0,2916 + 0,0486 = 0,9963$$

Obs. : $P(X < 3)$ pode ser obtida pelo seu complemento $P(X \geq 3)$:

$$P(X < 3) = 1 - P(X \geq 3) = 1 - 0,0037 = 0,9963$$

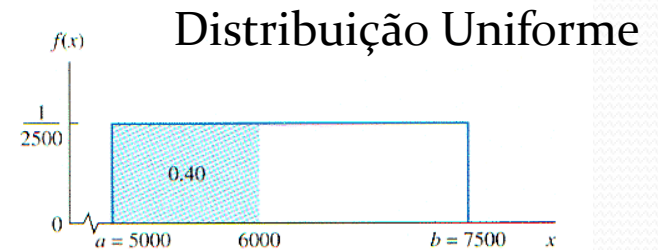
Distribuições Contínuas

- Uma função de densidade de probabilidade contínua corresponde à expressão matemática que define a distribuição dos valores para uma variável aleatória contínua.

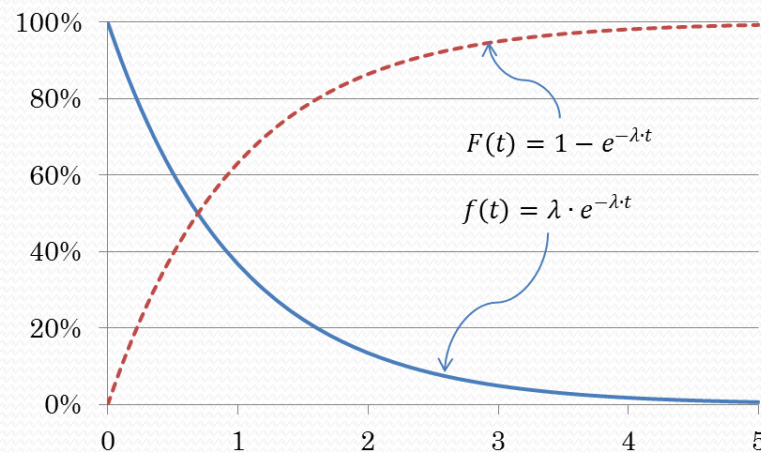


Distribuição Normal

FIGURE 8.1
Continuous uniform probability distribution with $a = 5000$, $b = 7500$ —Population example



Distribuição Uniforme



Distribuição Exponencial (azul)

Distribuição Normal

- Também chamada de *Distribuição de Gauss*.
- É uma distribuição contínua mais utilizada em estatística.
- É importante pois:
 - Inúmeras variáveis contínuas comuns no mundo dos negócios possuem distribuições que se assemelham estreitamente à normal.
 - Pode ser utilizada para fazer aproximações para várias distribuições de variáveis discretas.
 - A distribuição normal proporciona a base para a inferência estatística clássica, em razão de sua relação com o Teorema do Limite Central.
- Pode-se calcular a probabilidade de que vários valores ocorram dentro de determinadas amplitudes ou intervalos, mas a **probabilidade exata de um valor específico é zero**.

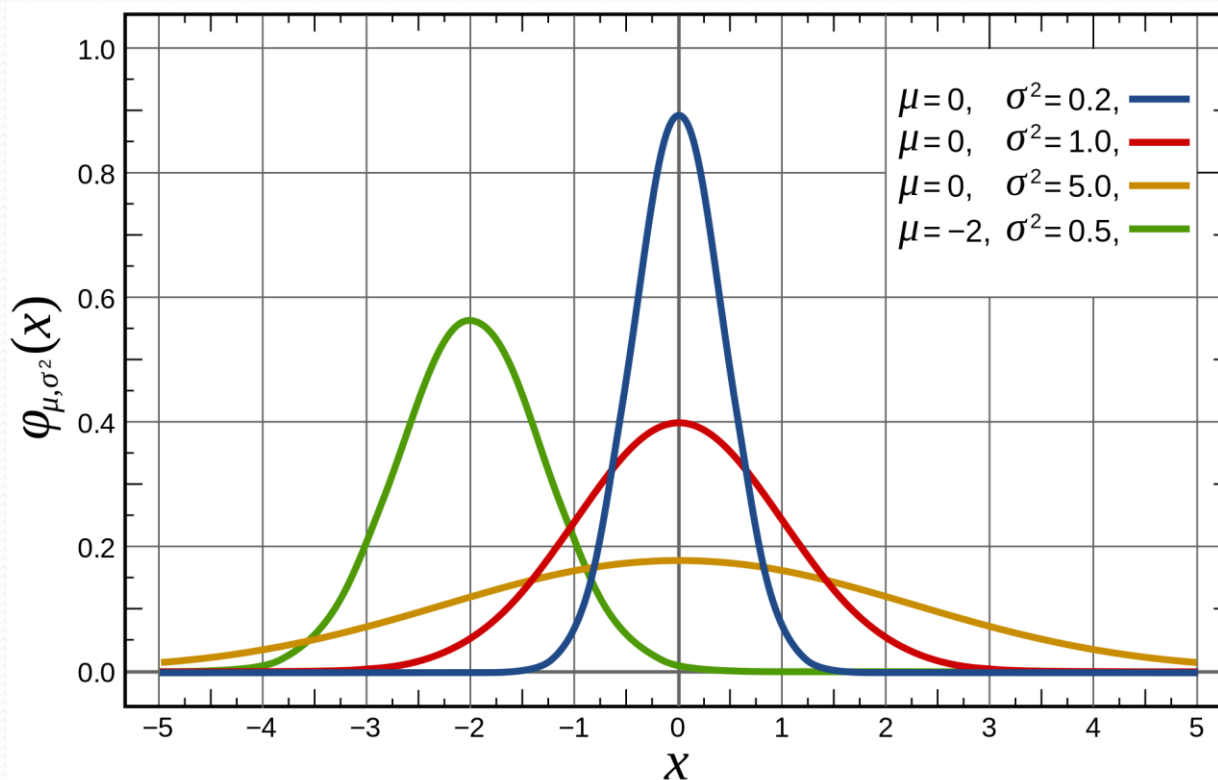
Distribuição Normal

- Propriedades:
 - Tem o formato de um sino (simétrica).
 - Tem média, mediana e moda são iguais.
 - A amplitude interquartil ($Q_3 - Q_1$) é igual a 1,33 desvio-padrão. Isso significa que os 50% valores centrais estão contidos dentro de um intervalo que tem como limites dois terços de um desvio-padrão abaixo da média e dois terços de um desvio-padrão acima da média.
 - Possui uma amplitude infinita ($-\infty < X < +\infty$).

Expressão Matemática

- A expressão matemática que representa a função de densidade da probabilidade normal é:

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



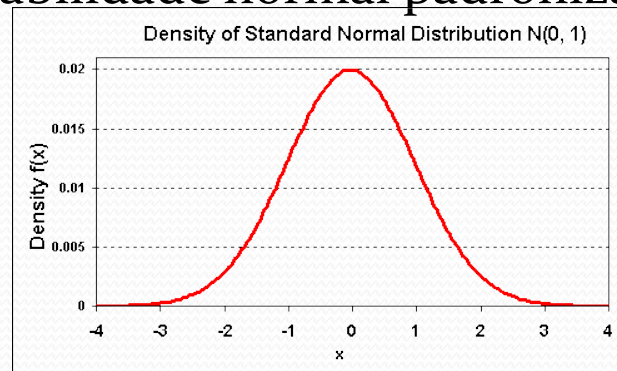
Função Padronizada

- A expressão vista é entediante em termos de cálculos e requer o cálculo de integrais.
- Existem tabelas para obtenção das probabilidades normais, mas apenas para $\mu=0$ e $\sigma=1$.
- Para converter uma média e um desvio-padrão qualquer para os valores normalizados ($\mu=0$ e $\sigma=1$) utilizamos a variável aleatória Z:

$$Z = \frac{X - \mu}{\sigma}$$

- Função de densidade da probabilidade normal padronizada

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



Exemplo de aplicação

- Está sendo desenvolvido uma pagina Web para a *OurCampus!*, que tem como público-alvo uma faculdade. Para atrair e reter usuários, é preciso garantir que a página inicial seja rapidamente aberta. Tanto o desenho da página inicial como a carga dos servidores de rede da empresa afetam o tempo de abertura.
- Para verificar a velocidade de abertura da página inicial, abre-se o navegador em um PC nos escritórios da *OurCampus!* e mede-se o tempo de abertura em segundos, desde o primeiro clique no link para o endereço na Web até que a página inicial esteja totalmente carregada.
- Dados do passado indicam que a **média aritmética** do tempo de abertura é de **7 segundos** e que o **desvio-padrão** é de **2 segundos**. Aproximadamente *dois terços* dos tempos estão entre *5 e 9 segundos*, e cerca de *95% dos tempos* de abertura estão entre *3 e 11 segundos*, ou seja, os tempos de abertura estão distribuídos como uma curva em *formato de sino*, com uma concentração em torno da média.

Variável Z

- Utilizando a expressão $Z = \frac{X - \mu}{\sigma}$ converta os seguintes valores de X para Z:
- X=1, X=3, X=5, X=7, X=9, X=11 e X=13

$$Z = \frac{X - \mu}{\sigma} \text{ onde } \mu = 7 \text{ e } \sigma = 2$$

$$Z = \frac{1-7}{2} = -3$$

$$Z = \frac{3-7}{2} = -2$$

$$Z = \frac{5-7}{2} = -1$$

$$Z = \frac{7-7}{2} = 0$$

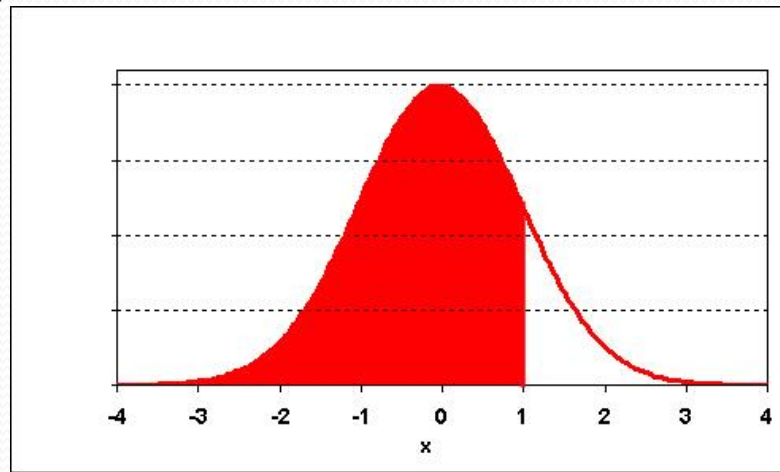
$$Z = \frac{9-7}{2} = +1$$

$$Z = \frac{11-7}{2} = +2$$

$$Z = \frac{13-7}{2} = +3$$

X (s)	Z
1	-3
3	-2
5	-1
7	0
9	1
11	2
13	3

- Encontre a probabilidade de que o tempo necessário para abertura da página da OurCampus! Seja menor que 9 segundos.
- Neste caso, como $X=9$, $Z=+1$.
- A probabilidade desejada pode ser obtida pela área sob a distribuição para Z menor do que +1:



- Para se encontrar essa área, devemos calcular a integral sob a curva entre os limites $-\infty$ e $+1$.
- Essa integral pode ser calculada manualmente, numericamente (computador) ou usando tabelas.

$$P(Z \leq +1) = \int_{-\infty}^{+1} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = \int_{-\infty}^{<+1} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz + \int_{+1}^{+1} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz =$$

$$= P(Z < +1) + 0 = P(Z < +1)$$

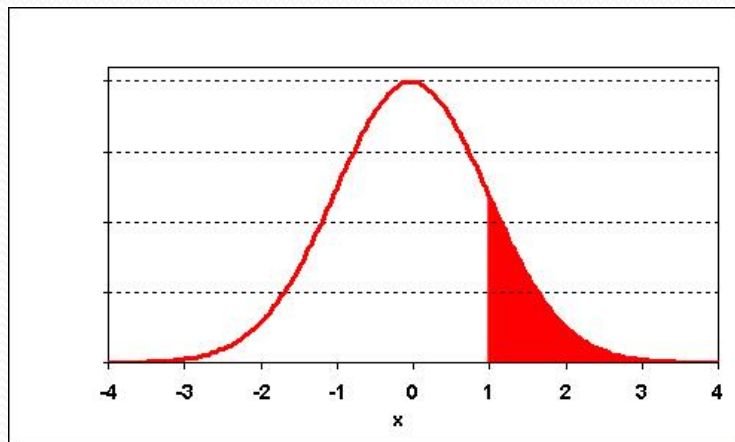
- Usando a tabela (probabilidade entre $-\infty$ e Z):

Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7518	0,7549
0,7	0,7580	0,7612	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621

Fonte: Extraída da Tabela E.2.

- Assim: $P(Z < +1) = 0,8413$
- A probabilidade do tempo de abertura ser *menor do que 9 segundos é de 0,8413*.

- Qual é a probabilidade de que o tempo de abertura venha a ser maior que 9 segundos?

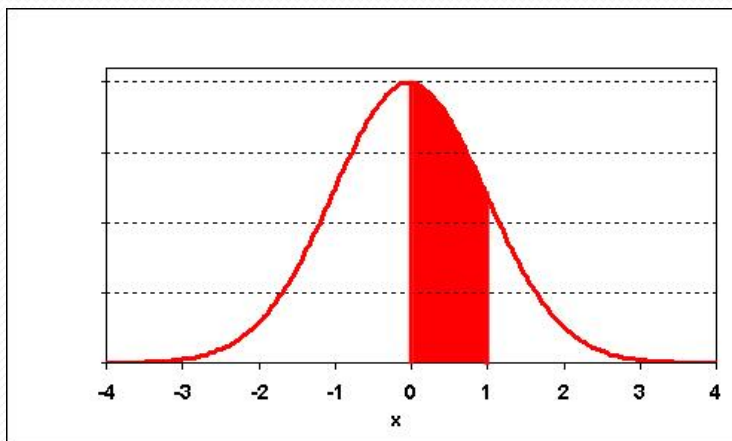


- A probabilidade de que o tempo de abertura venha a ser menor do que 9 segundos é 0,8413 (já calculada). Assim, a probabilidade de que o tempo de abertura venha a ser *maior do que 9 segundos* é igual ao **complemento** de menos do que 9 segundos:

$$P(X > 9) = 1 - P(X < 9) = 1 - 0,8413 = 0,1587$$

- Assim, a probabilidade de o tempo de abertura ser maior do que 9 segundos é de 0,1587.

- Qual é a probabilidade de que o tempo de abertura venha a estar entre 7 e 9 segundos?
- Neste caso, queremos $P(Z > 0 \text{ e } Z < +1)$.



- A probabilidade do tempo ser menor do que 9 segundos é 0,8413.
- A probabilidade desejada pode ser obtida:

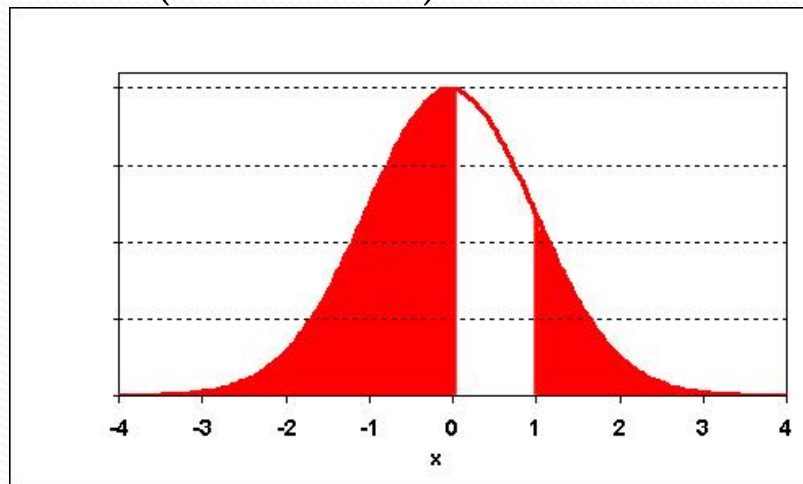
$$P(Z > 0 \text{ e } Z < +1) = P(Z < +1) - P(Z < 0)$$

Mas $P(Z < 0) = 0,5000$, então :

$$P(Z > 0 \text{ e } Z < +1) = P(Z < +1) - 0,5000 = 0,8413 - 0,5000 = 0,3413$$

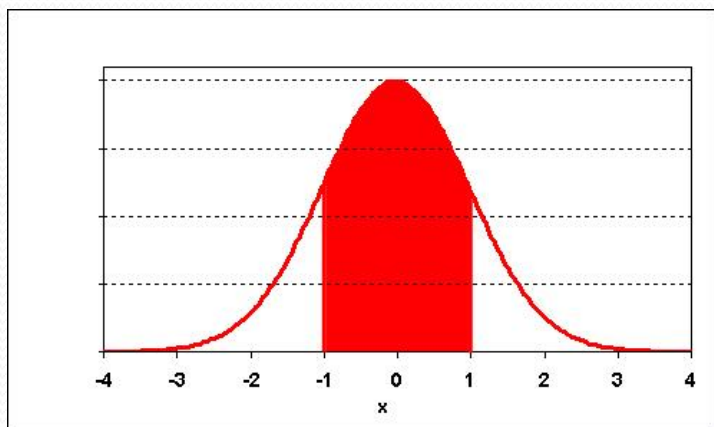
- Assim, a probabilidade de o tempo de abertura estar entre 7 e 9 segundos é de 0,3413.

- Qual é a probabilidade de que o tempo de abertura seja menor do que 7 segundos ou maior do que 9 segundos?
- Neste caso, queremos $P(Z < 0 \text{ ou } Z > 1)$.



- A probabilidade procurada é o complemento de $P(Z > 0 \text{ e } Z < +1)$:
$$P(Z < 0 \text{ ou } Z > +1) = 1 - P(Z > 0 \text{ e } Z < +1) = 1 - 0,3413 = 0,6587.$$
- Assim, a probabilidade de que o tempo de abertura seja menor do que 7 segundos e maior do que 9 segundos é de 0,6587.

- Qual a probabilidade que o tempo de abertura venha a estar entre 5 e 9 segundos?
- Queremos obter $P(Z > -1 \text{ e } Z < +1)$.



$$P(Z > -1 \text{ e } Z < +1) = P(Z < +1) - P(Z < -1)$$

Da tabela, $P(Z < -1) = 0,1587$

Assim :

$$P(Z > -1 \text{ e } Z < +1) = 0,8413 - 0,1587 = 0,6826$$

- Essa probabilidade corresponde ao intervalo definido pelo desvio-padrão ($\sigma=1$ na padronizada, $\sigma=2$ no exemplo dado).
- Assim, a probabilidade de o tempo de abertura estar entre 5 e 9 ou de estar entre $\mu-\sigma$ e $\mu+\sigma$ é de 0,6826.
- O desvio-padrão define um intervalo em torno de média de probabilidade de 68,26%.

- Encontre as probabilidades dos tempos de abertura venha a estar entre 3 e 11 segundos e entre 1 e 13 segundos.

- Este problema tem solução igual ao exemplo anterior:

$$P(X > 3 \text{ e } X < 11) = P(Z > -2 \text{ e } Z < +2) = P(Z < +2) - P(Z < -2)$$

$$\text{Da tabela : } P(Z < +2) = 0,9772 \text{ e } P(Z < -2) = 0,0228$$

$$P(Z < +2) - P(Z < -2) = 0,9772 - 0,0228 = 0,9544$$

$$P(X > 1 \text{ e } X < 13) = P(Z > -1 \text{ e } Z < +1) = P(Z < +1) - P(Z < -1)$$

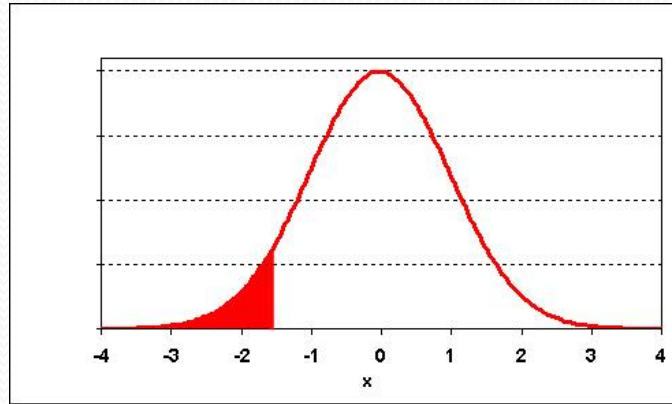
$$\text{Da tabela : } P(Z < +3) = 0,99865 \text{ e } P(Z < -3) = 0,00135$$

$$P(Z < +3) - P(Z < -3) = 0,99865 - 0,00135 = 0,9973$$

- Resumindo:

- Aproximadamente 68,26% dos valores se posicionam dentro dos limites de +/-1 desvio-padrão de distância em relação à média.
- Aproximadamente 95,44% dos valores se posicionam dentro dos limites de +/-2 desvios-padrão de distância em relação à média.
- Aproximadamente 99,73% dos valores se posicionam dentro dos limites de +/-3 desvios-padrão de distância em relação à média.

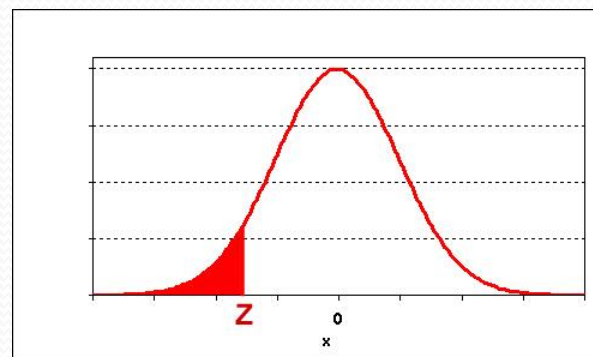
- Qual é a probabilidade de que um tempo de abertura venha a ser inferior a 3,5 segundos?
- Neste caso, queremos $P(Z < -1,75)$ - calcular).



$$P(Z < -1,75) = 0,0401 \text{ da tabela}$$

- Assim, a probabilidade de o tempo de abertura ser menor do que 3,5 segundos é de 0,0401.

- Quanto tempo (em segundos) terá decorrido antes que 10% das aberturas de páginas tenham sido completadas?
- Uma vez que é esperado que 10% das páginas da Web sejam abertas em menos do que X segundos, queremos obter a área sob a curva normal que corresponde a 0,1000.
- Devemos procurar na tabela o valor de Z que corresponde a uma área de 0,1000:



Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
-1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
-1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
-1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
-1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985

Fonte: Extraída da Tabela E.2.

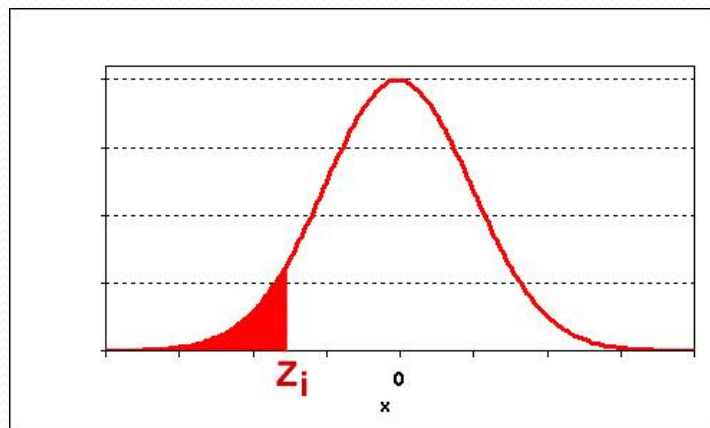
- O valor de Z mais próximo do desejado é Z=-1,28:

$$Z = \frac{X - \mu}{\sigma} \Rightarrow X = \mu + Z\sigma \quad \text{Como } Z = -1,28, \mu = 7 \text{ e } \sigma = 2 :$$

$$X = 7 - 1,28 \times 2 = 4,44 \text{ segundos}$$

- Assim, 10% dos tempos de abertura correspondem a 4,44 segundos ou menos.

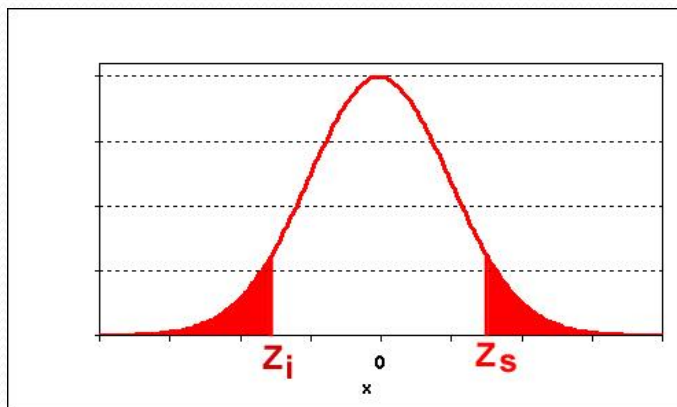
- Quais são os valores superior e inferior de X , distribuídos simetricamente em torno da média, que incluem 95% dos tempos de abertura?
- Inicialmente, precisamos encontrar o valor *inferior* de Z (Z_i e posteriormente X_i). Depois, devemos encontrar o valor *superior* de Z (Z_s e posteriormente X_s).
- Uma vez que 95% dos valores estão entre Z_i e Z_s (X_i e X_s) e que se encontram equidistantes da média, a área sob $Z < Z_i$ e $Z > Z_s$ devem ser *iguais* a 2,50% cada uma (0,0250).



- Na tabela, a área inferior desejada (0,0250) é determinada por $Z = -1,96$:

Como $Z_i = -1,96$, obtemos $X_i = \mu + Z\sigma = 7 - 1,96 \times 2 = 3,08$ segundos

- De forma semelhante, devemos determinar Z_s :



- A tabela nos fornece a área para $Z < Z_s$. Como queremos a área de $Z > Z_s$ que vale 0,025, a área que devemos procurar na tabela é $1 - 0,025 = 0,9750$:

Pela tabela, $Z_s = +1.96$

Então $X = \mu + Z\sigma = 7 + 1,96 \times 2 = 10,92$ segundos

- Portanto, o intervalo desejado de 95% dos tempos de abertura está compreendido entre 3,08 segundos e 10,92 segundos.

Bibliografia

- Levine, Stephan, Krehbiel, Berenson. Estatística, Teoria e Aplicações usando o Mirosoft Excel em Português. Quinta Edição. Ed. LTC.