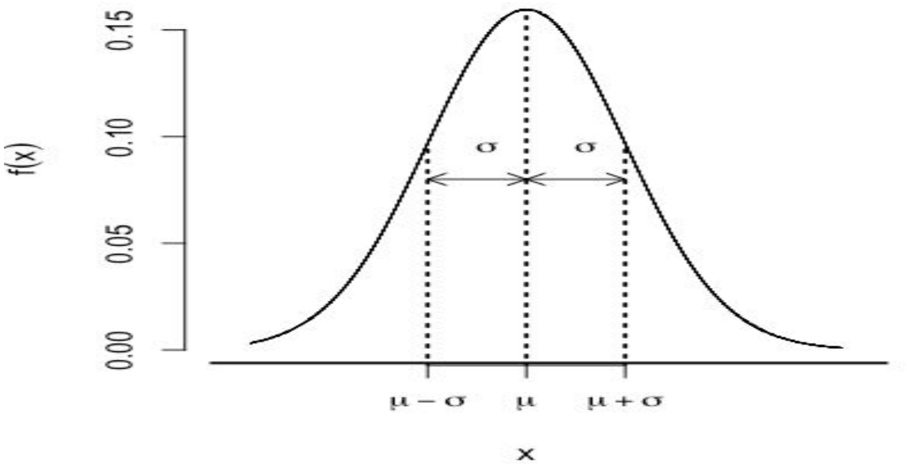
DISTRIBUIÇÃO NORMAL

Uma v.a. X tem distribuição normal com parâmetros μ ∈ R e σ >0 se sua função densidade de probabilidade é dada por

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

Notação: $X \sim N(\mu, \sigma^2)$

Função densidade de probabilidade



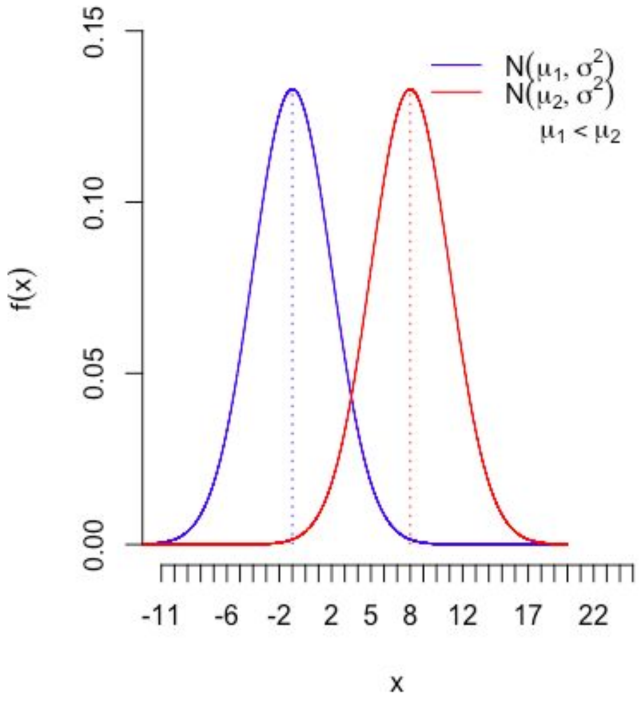
Função distribuição acumulada

Não tem expressão fechada.

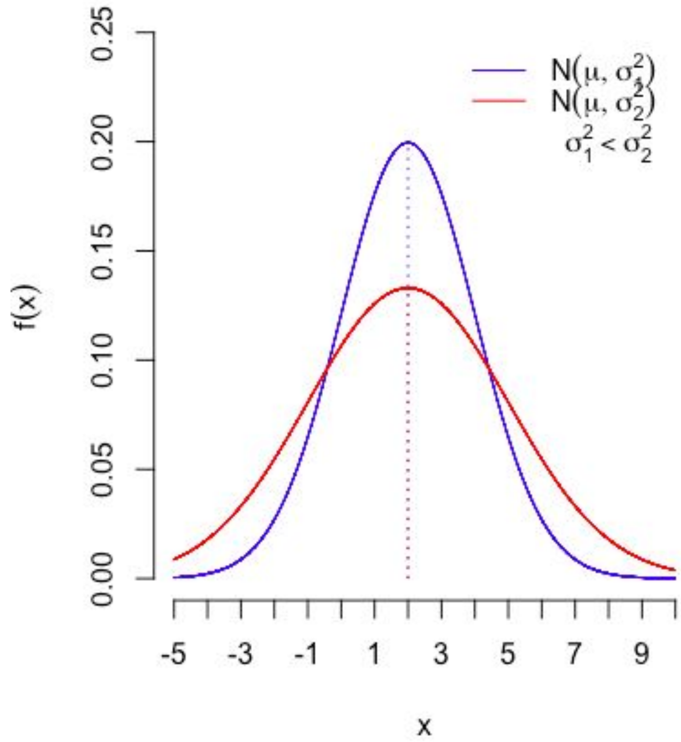
Valor esperado e variância

$$E(X) = \mu \text{ e } V(X) = \sigma^2$$

Médias diferentes e mesma variância



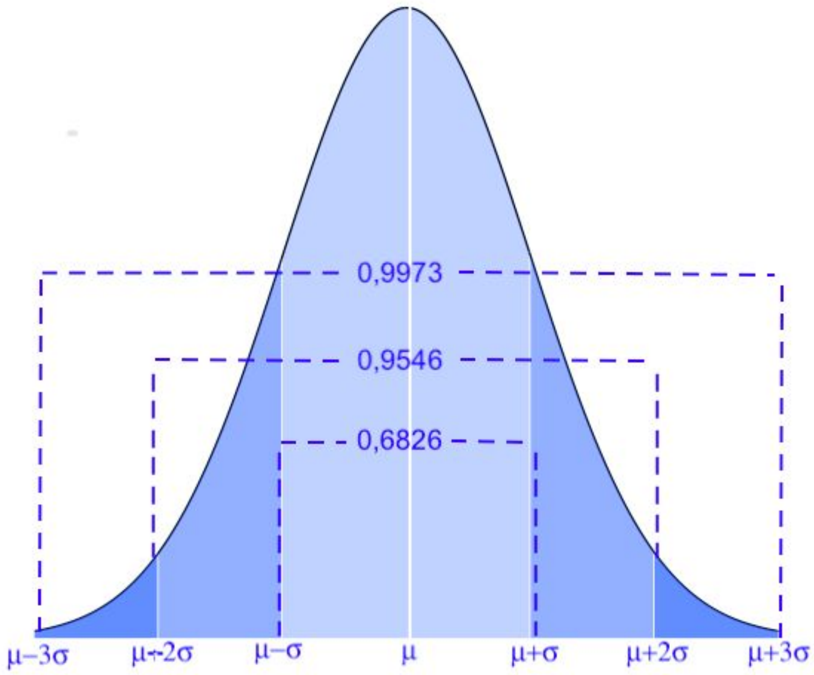
Médias iguais e variâncias diferentes



CARACTERÍSTICAS DA DISTRIBUIÇÃO NORMAL

- É **simétrica** em relação à média.
- Média, mediana e moda são iguais.
- A área sob a curva de densidade à esquerda da média é igual a 0,5.
- A área sob a curva de densidade à direita da média é igual a 0,5.
- Quando a média é zero e a variância é 1, é chamada de **normal padrão**.
- O parâmetro μ (média) é chamado de parâmetro de **locação**.
- O parâmetro σ (desvio padrão) é chamado de parâmetro de **escala**.

Área sob a curva



Transformação linear de uma variável normal

Se $X \sim N(\mu,\sigma^2)$ então $Y = a + bX \sim N(a+ b\mu, b^2\sigma^2)$.

Caso particular: Se Z é uma transformação linear de X tal que $a = -\mu/\sigma$ e $b = 1/\sigma$ então

$$Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$

Tabela da Distribuição Normal

A fda da N(0,1) é tabelada, ou seja, existe uma tabela que contém a aproximação da fda para diferentes quantis (z).

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}\, dt$$

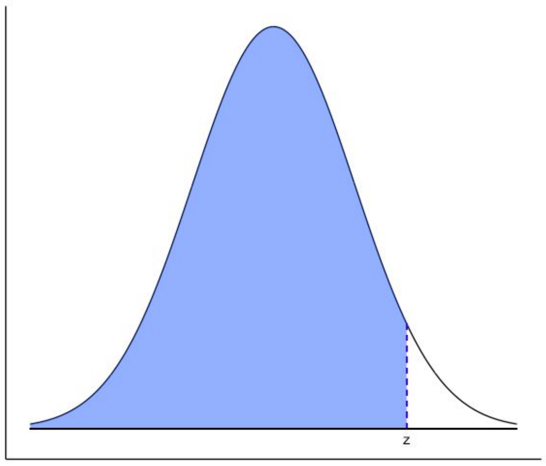
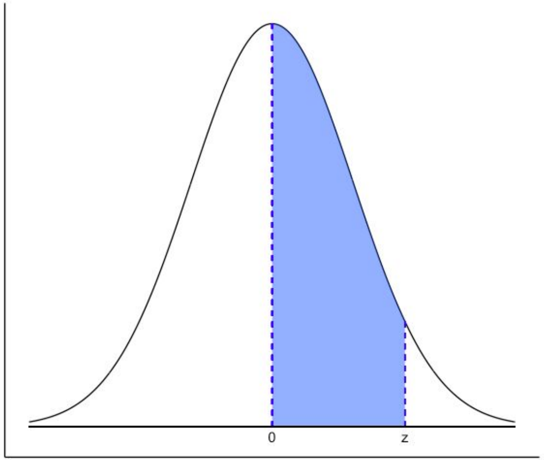


Tabela da Distribuição Normal

$$P(0 \le Z \le z)$$



$$P(Z > z)$$

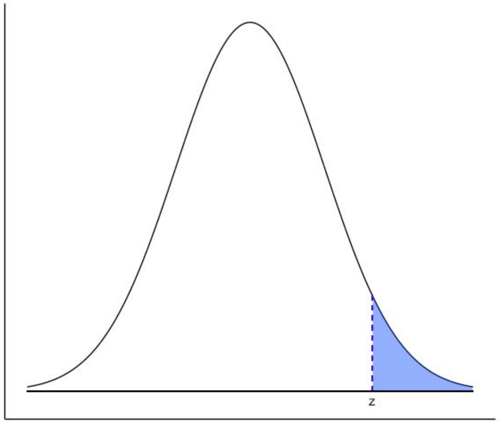


Tabela da Distribuição Normal (Padrão)

$P(Z<z)$

z	0,0	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	0,7517	0,7549
0,7	0,7580	0,7611	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
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

Tabela da Distribuição Normal (Padrão)

P(Z≤-0,65)

P(Z<z)

Segunda decimal

Parte inteira e primeira decimal

z	0,0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
-0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
-0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
-0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
-0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
-0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
-0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
-0,7	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
-0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
-0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
-1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
-1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
-1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
-1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823

Usando um software

Quando vamos usar um software devemos buscar como resposta o quantil, quando queremos saber z, e como ter resposta a probabilidade, quando queremos saber $P(Z \leq z)$.

Não precisamos padronizar, basta sabermos se o software pede como informação o desvio padrão, a variância ou a precisão da distribuição normal, e informar esta medida junto com a média.

Exemplo

Calcular: $P(X \leq x) = p$, sendo $X \sim N(\mu,\sigma^2)$

No R:

p = pnorm(x, μ, σ)

x = qnorm(p, μ, σ)

No Excel:

p = DIST.NORM.N(x; μ; σ; VERDADEIRO)

x = INV.NORM.N(p; μ; σ)

Exemplo

O tempo necessário para produzir um lote de itens tem distribuição normal com média 120 minutos e desvio padrão 15 minutos.

(a) Sorteando-se um lote produzido, qual a probabilidade de que tempo de produção seja inferior a 100 minutos?

(b) Qual o tempo correspondente à produção de 95% dos itens?

(c) Qual o intervalo de tempo central correspondente à produção de 80% dos itens?

X: tempo de prod. $\quad X \sim N(120, 15^2)$

Solução

$$P(X < 100) = P\Big(\underbrace{\frac{X-120}{15}}_{Z \sim N(0,1)} < \frac{100-120}{15}\Big)$$

$$= P(Z < -1{,}33) = 0{,}0918$$

$$\text{pnorm}(-1.33) = 0{,}0912$$

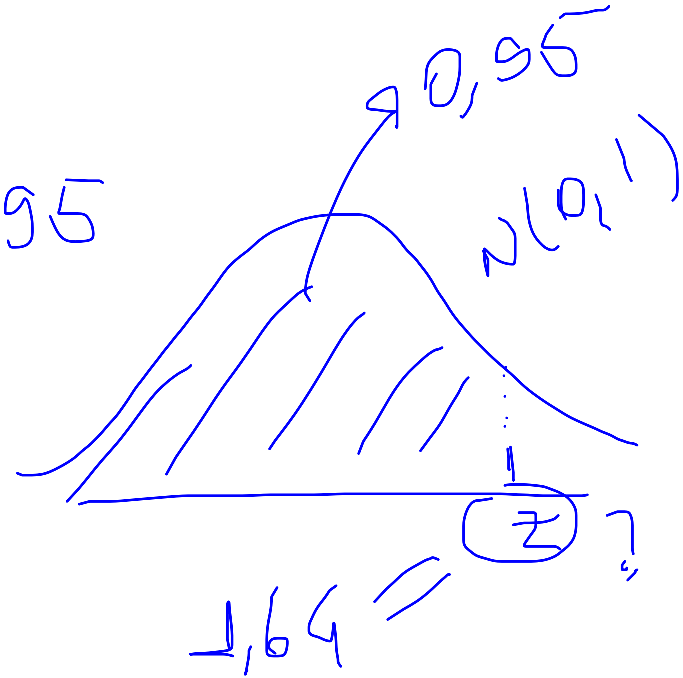
R $\quad P(X < 100) = \text{pnorm}(100, 120, 15) = 0{,}0912$

Solução

$$P(X < x) = 0,95$$

$$P\left(\frac{X-120}{15} < \frac{x-120}{15}\right) = 0,95$$

$$P\Big(Z < \underbrace{\frac{x-120}{15}}_{1,64}\Big) = 0,95$$



$$\frac{x-120}{15} = 1,64 \Rightarrow x = 144,6 \text{ min.}$$

qnorm(0.95, 120, 15) = 144,67

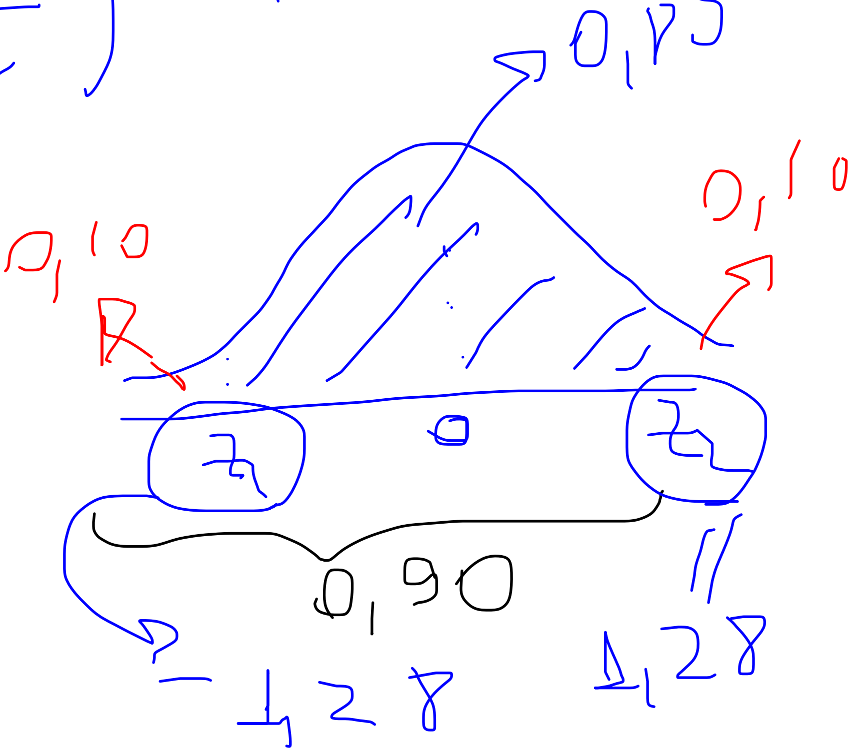
Solução

$$P(x_1 < X < x_2) = 0{,}80$$

$$P\Big(\underbrace{\frac{x_1-120}{15}}_{z_1} < Z < \underbrace{\frac{x_2-120}{15}}_{z_2}\Big) = 0{,}80$$

$$\frac{x_1-120}{15} = -1{,}28 \Rightarrow x_1 = 100{,}8 \text{ min}$$

$$\frac{x_2-120}{15} = 1{,}28 \Rightarrow x_2 = 139{,}2 \text{ min}$$



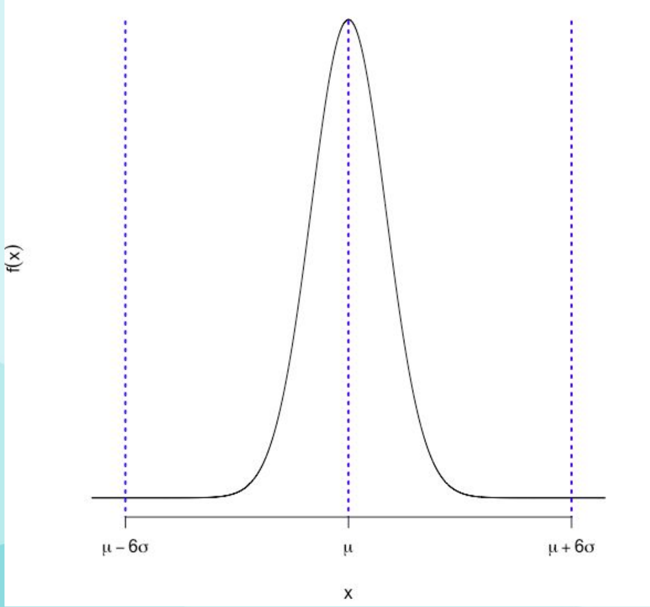
Resp: $(100{,}8 ; 139{,}2)$

A ESCALA SIGMA

Utilizada para medir o nível de **qualidade** de um processo de produção. Quanto **maior** o número de **sigmas** (σ), **melhor**.

A ESCALA SIGMA

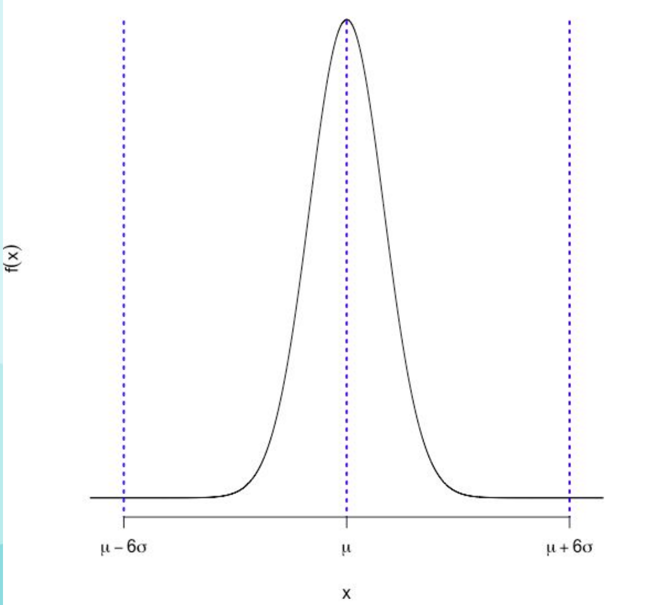
- X representa uma característica de um item, sendo que $X \sim N(\mu,\sigma^2)$, sendo μ o valor nominal.



- Limites de especificação: LIE = μ - 6σ e LSE = μ + 6σ. Então:
- $$P(X < \mu - 6\sigma) + P(X > \mu + 6\sigma) = 1,973175 \times 10^{-9}$$
- Corresponde, em **média**, a cerca de **dois** itens que não atendem às especificações a cada **bilhão** de itens produzidos.

2 a cada bilhão

A ESCALA SIGMA



- O processo sofre uma alteração.
- A média para a ser $\mu_A = \mu + 1{,}5\sigma$, ou seja, a nova produção segue $X \sim N(\mu_A, \sigma^2)$. Então
- $$P(X < \mu - 6\sigma) + P(X > \mu + 6\sigma) = 3,397673 \times 10^{-6}$$
- Corresponde, em **média**, a cerca de **3,4** itens que não atendem às especificações a cada **milhão** de itens produzidos.

3,4 a cada milhão

A ESCALA SIGMA

Nível	Média de defeitos por milhão
2σ	308537
3σ	66807
4σ	6210
5σ	233
6σ	3,4

A ESCALA SIGMA

4σ	6σ
Sete horas de falta de energia por **mês**	**Uma hora** de falta de energia a cada **20 anos**
3,2 cirurgias incorretas por **semana** (500/semana)	**3,4** cirurgias incorretas a cada **5 meses**
15 minutos de fornecimento de água não potável por **dia**	**Um minuto** de fornecimento de água não potável a cada **sete meses**

Fonte: Keene, S. (2000), *Reliability Review* 20, p.19.

APROXIMAÇÃO DA BINOMIAL PELA NORMAL

Se X é uma v.a. com distribuição binomial, Bin(p,n), de média E(X) = np e variância V(X) = np(1-p), então para n suficientemente grande ($n \to \infty$) temos que

$$Z = \frac{X - np}{\sqrt{np(1-p)}} \sim \mathcal{N}(0,1)$$

QUAL A VANTAGEM?

$$P(X \leq x) = \sum_{k=0}^{x} \binom{n}{k} p^k (1-p)^{1-k}$$

Usando a aproximação pela normal temos

$$P\left(\frac{X - np}{\sqrt{np(1-p))}} \leq \frac{x - np}{\sqrt{np(1-p)}}\right) = P\left(Z \leq \frac{x - np}{\sqrt{np(1-p)}}\right)$$

APROXIMAÇÃO NORMAL E PROBABILIDADES BINOMIAIS VERDADEIRAS

	$p = 0{,}05$, $n = 10$		$p = 0{,}10$, $n = 10$		$p = 0{,}50$, $n = 10$	
r	**Binomial**	**Normal**	**Binomial**	**Normal**	**Binomial**	**Normal**
0	0,5987	0,5000	0,3487	0,2981	0,0010	0,0022
1	0,9139	0,9265	0,7361	0,7019	0,0107	0,0136
2	0,9885	0,9981	0,9298	0,9429	0,0547	0,0571
3	0,9990	1,0000	0,9872	0,9959	0,1719	0,1711
4	1,0000	1,0000	0,9984	0,9999	0,3770	0,3745
5			1,0000	1,0000	0,6230	0,6255
6					0,8281	0,8289
7					0,9453	0,9429
8					0,9893	0,9864
9					0,9990	0,9978
10					1,0000	0,9997

(*continua*)

Fonte: Walpole, Ronald E. *Probabilidade & Estatística para Engenharia e Ciências*. Pearson Prentice Hall, 2009.

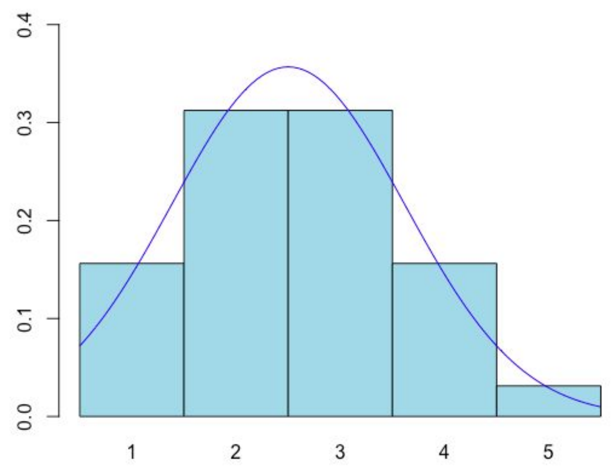
APROXIMAÇÃO NORMAL E PROBABILIDADES BINOMIAIS VERDADEIRAS

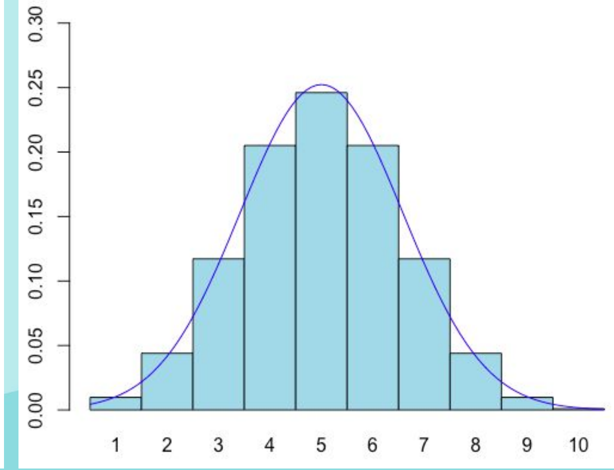
(continuação)

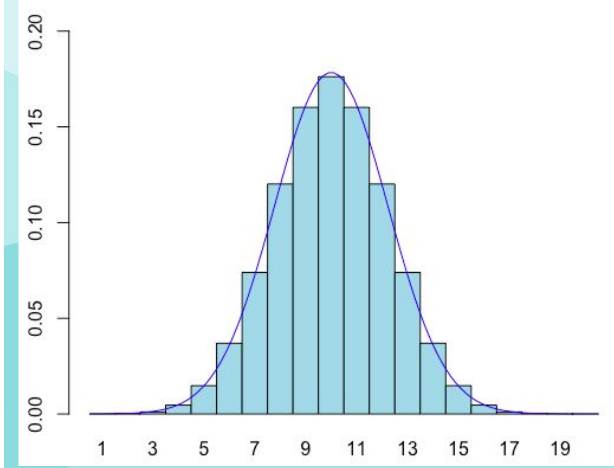
	$p = 0{,}05$					
	$n = 20$		$n = 50$		$n = 100$	
r	**Binomial**	**Normal**	**Binomial**	**Normal**	**Binomial**	**Normal**
0	0,3585	0,3015	0,0769	0,0968	0,0059	0,0197
1	0,7358	0,6985	0,2794	0,2578	0,0371	0,0537
2	0,9245	0,9382	0,5405	0,5000	0,1183	0,1251
3	0,9841	0,9948	0,7604	0,7422	0,2578	0,2451
4	0,9974	0,9998	0,8964	0,9032	0,4360	0,4090
5	0,9997	1,0000	0,9622	0,9744	0,6160	0,5910
6	1,0000	1,0000	0,9882	0,9953	0,7660	0,7549
7			0,9968	0,9994	0,8720	0,8749
8			0,9992	0,9999	0,9369	0,9463
9			0,9998	1,0000	0,9718	0,9803
10			1,0000	1,0000	0,9885	0,9941

Fonte: Walpole, Ronald E. *Probabilidade & Estatística para Engenharia e Ciências*. Pearson Prentice Hall, 2009.

Comparação para p = 0,5





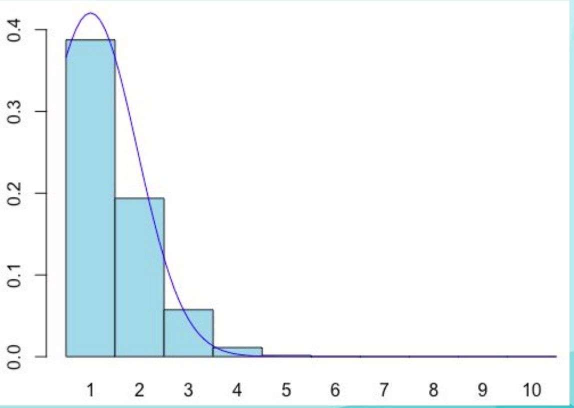


n=5

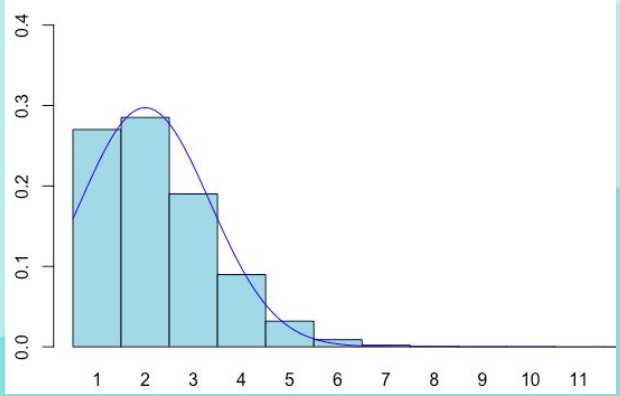
n=10

n=20

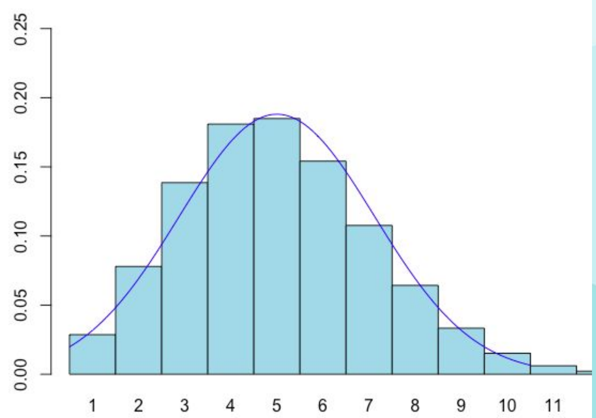
Comparação para p = 0,1



n=10



n=20



n=50

Exemplo

O RH de uma empresa recebeu 100 currículos de candidatos. A probabilidade de um candidato preencher os requisitos para a vaga é de 0,20. Se existem 12 vagas na empresa, qual a probabilidade de todas as vagas serem preenchidas?

Solução

X : nº bons candidatos

$X \sim Bin(100; 0,2)$

$$\underbrace{P(X \geq 12)}_{} = \sum_{x=12}^{100} \binom{100}{x} 0,2^{x} (1-0,2)^{100-x}$$

$$= 1 - \sum_{x=0}^{11} \binom{100}{x} 0,2^{x} (1-0,2)^{100-x}$$

$$\simeq P\left(\underbrace{\frac{X - 100 \cdot 0,2}{\sqrt{100 \cdot 0,2 \cdot 0,8}}}_{Z \sim N(0,1)} \geq \frac{12 - 100 \cdot 0,2}{\sqrt{100 \cdot 0,2 \cdot 0,8}} \right)$$

$$P(Z \geq -2) = 1 - P(Z < -2) = \boxed{0,9772}$$