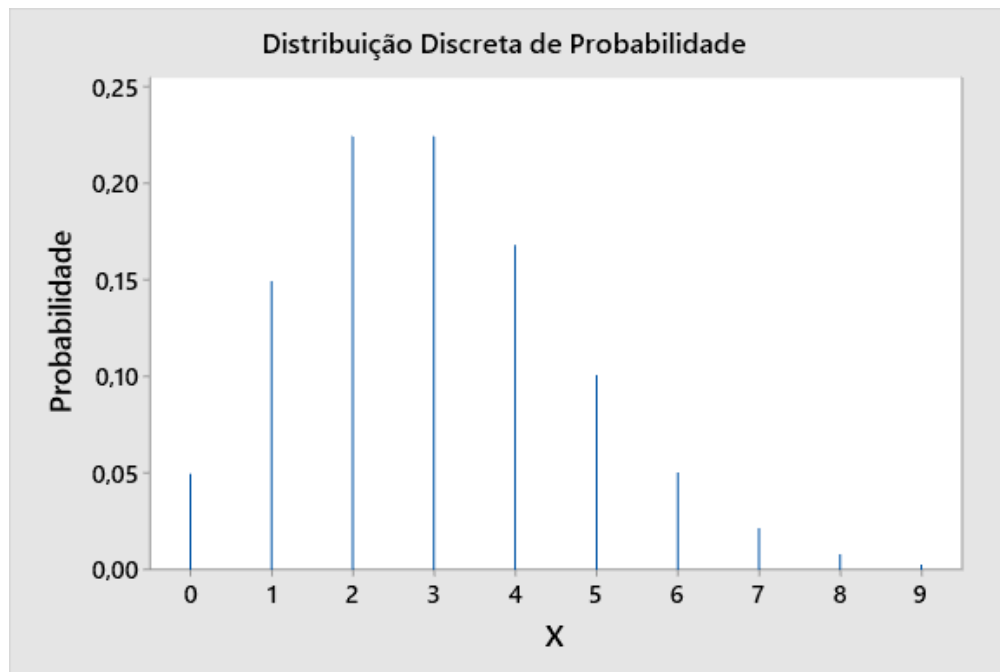


Resolução de Exercícios

Profª Julienne Borges

Distribuição discreta de probabilidade



Lembrete!

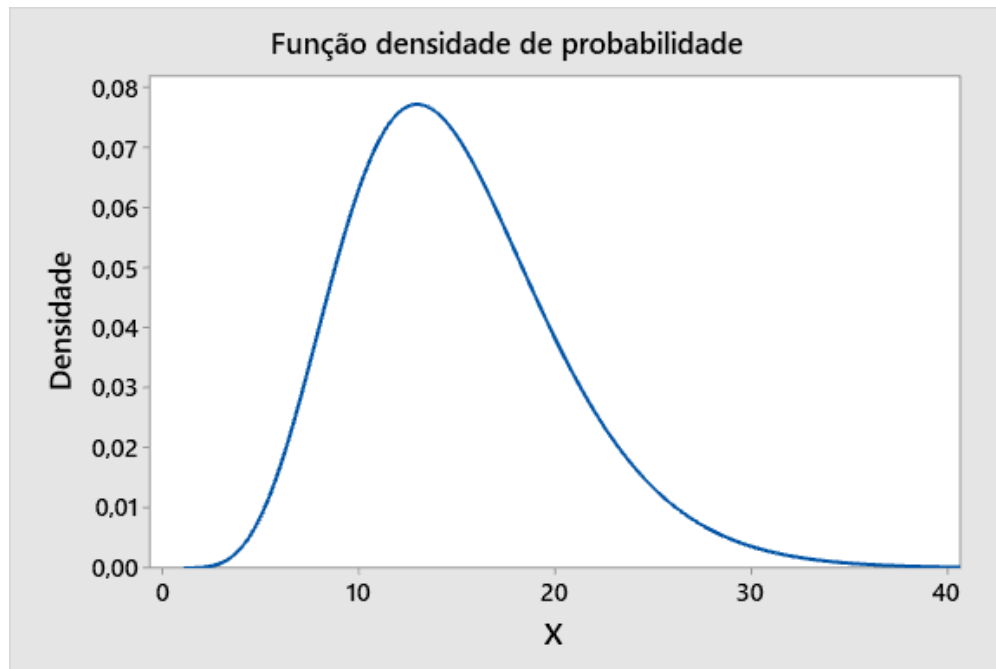
$$P(X < 1) = P(X = 0)$$

$$P(X \leq 1) = P(X = 0) + P(X = 1)$$

$$P(X > 7) = P(X = 8) + P(X = 9)$$

$$P(X \geq 7) = P(X = 7) + P(X = 8) + P(X = 9)$$

Função densidade de probabilidade



Lembrete!

$$P(X < 10) = P(X \leq 10)$$

$$P(X > 20) = P(X \geq 20)$$

Assim como

$$\begin{aligned} P(10 < X < 20) &= P(10 \leq X \leq 20) = \\ &= P(10 < X \leq 20) = P(10 \leq X < 20) = \end{aligned}$$

Distribuição Normal

Como saber se meus dados têm distribuição normal?

Opção 1:

Selecionar uma amostra, construir o histograma para os dados e verificar se a forma do gráfico é simétrica.

É a opção menos indicada por ser muito susceptível a erro!

Como saber se meus dados têm distribuição normal?

Opção 2:

O **gráfico quantil-quantil ou qq-plot**, é um dispositivo gráfico exploratório utilizado para verificar a validade de um pressuposto de distribuição para um conjunto de dados. Em geral, a ideia básica é a de calcular o valor teoricamente esperado para cada ponto de dados com base na distribuição em questão. Se os dados de fato seguirem a distribuição assumida os pontos deste gráfico formarão aproximadamente uma linha reta.

Em alguns casos você pode ficar na dúvida se realmente os dados seguem a distribuição avaliada ou não.

Como saber se meus dados têm distribuição normal?

Opção 3:

Realizar um teste de normalidade para os dados da sua amostra! Os testes de normalidade são usados para determinar se um conjunto de dados de uma dada variável aleatória, é bem modelada por uma distribuição normal ou não, ou para calcular a probabilidade da variável aleatória subjacente estar normalmente distribuída. Alguns possíveis testes de normalidade são: Teste de Shapiro-Wilk, Teste de Kolmogorov-Smirnov, Teste de Anderson-Darling, entre outros.

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5915855/>

<https://medium.com/dados-e-saude/testes-de-normalidade-9dc9068bcf80>

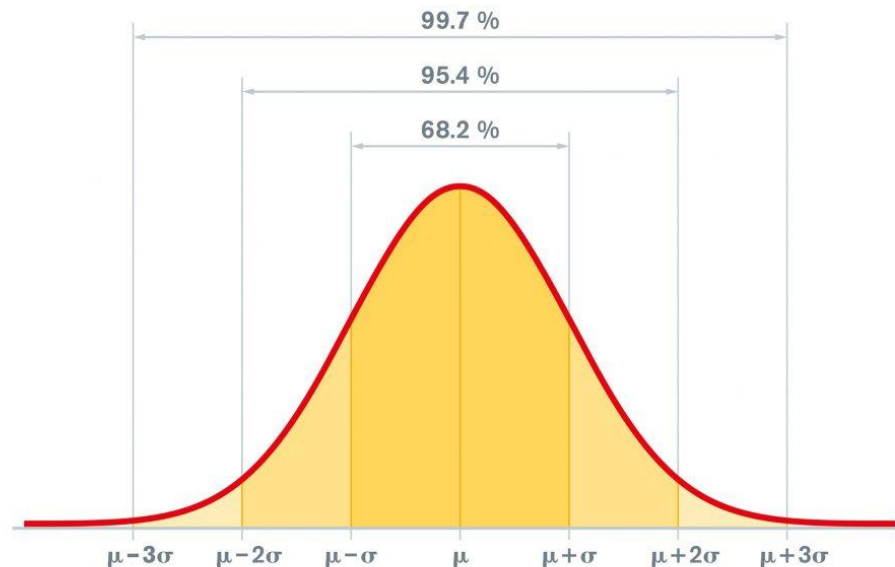
<https://blog.curso-r.com/posts/2021-01-28-testes-de-normalidade/>

<https://blog.curso-r.com/posts/2021-02-04-testes-de-normalidade/>

Regra empírica ou regra dos 3 sigmas ou regra 68-95-99

‘Meus dados têm distribuição normal com média 10 e desvio padrão igual a 2.’

Utilizando a regra empírica, sabemos que 99,7% dos dados estão no intervalo de 4 a 16!



Fonte: <https://datascience.eu/pt/matematica-e-estatistica/calculadora-de-regra-empirica/>

Utilizando o Microsoft Excel - Probabilidade

Distribuição Normal para o cálculo de probabilidade

$P(X \leq x)$

=DIST.NORM.N(x;media;desvio padrão;VERDADEIRO)

Utilizando o R - Probabilidade

Distribuição Normal para o cálculo de probabilidade

$$P(X \leq x)$$

`pnorm(x, m, s, lower.tail = TRUE)`

onde:

m= média

s= desvio padrão

lower.tail=TRUE → Fornece a probabilidade acumulada até x → $P(X \leq x)$

lower.tail=FALSE → Fornece a probabilidade acima de x → $P(X > x)$

Utilizando o Python - Probabilidade

Distribuição Normal para o cálculo de probabilidade

$P(X \leq x)$

`norm.cdf(x, m, s)`

Distribuição Normal para o cálculo de probabilidade

$P(X > x)$

`norm.sf(x, m, s)`

onde:

m= média

s= desvio padrão

Importante:

Para usar as funções de cálculo de probabilidade para a distribuição normal no Python é necessário primeiramente que você importe a função norm:

```
from scipy.stats import  
norm
```

Exercício 1

O porto de South Louisiana, localizado ao longo de 54 milhas do Rio Mississippi, entre Nova Orleans e Baton Rouge, é o maior porto de carga a granel do mundo. A quantidade de toneladas de carga movimentadas por semana é normalmente distribuída com uma média de 4,5 milhões de toneladas de carga por semana e um desvio padrão de 0,82 milhão de toneladas.



Exercício 1

X: Quantidade de carga movimentada por semana (em milhão de toneladas)



Distribuição Normal

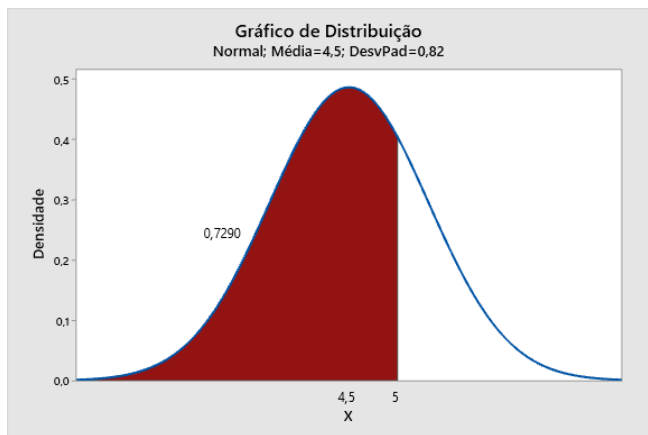


Parâmetros: $\mu = 4,5$ e $\sigma = 0,82$.



Exercício 1

- a) Qual é a probabilidade de o porto manejar menos de 5 milhões de toneladas de carga por semana?



$P(X < 5) = ?$

No Excel:

`=DIST.NORM.N(5;4,5;0,82;VERDADEIRO)`

No R:

`pnorm(5, 4.5, 0.82, lower.tail = TRUE)`

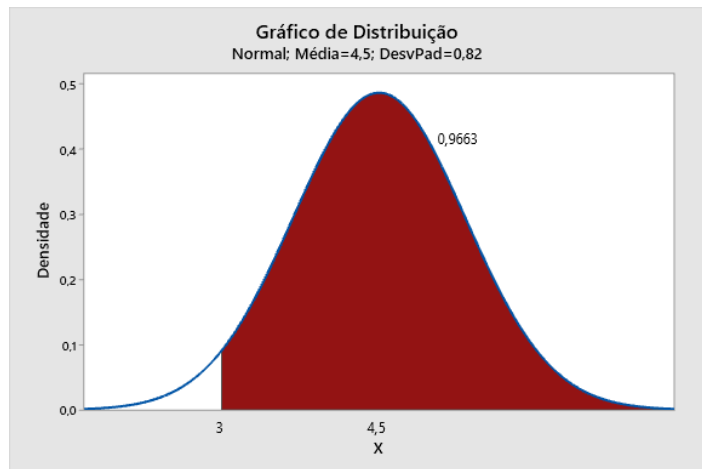
No Python:

`norm.cdf(5, 4.5, 0.82)`

Resultado: **0,7289883**

Exercício 1

b) Qual é a probabilidade de o porto movimentar 3 ou mais milhões de toneladas de carga por semana?



$P(X \geq 3) = ?$

No Excel: $P(X \geq 3) = 1 - P(X < 3)$
`=DIST.NORM.N(3;4,5;0,82;VERDADEIRO)`

No R:

`pnorm(3, 4.5, 0.82, lower.tail = FALSE)`

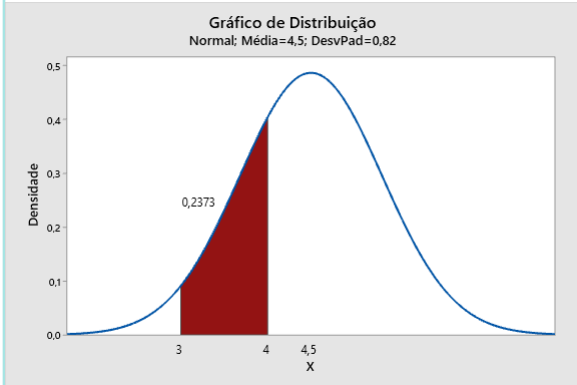
No Python:

`norm.sf(3, 4.5, 0.82)`

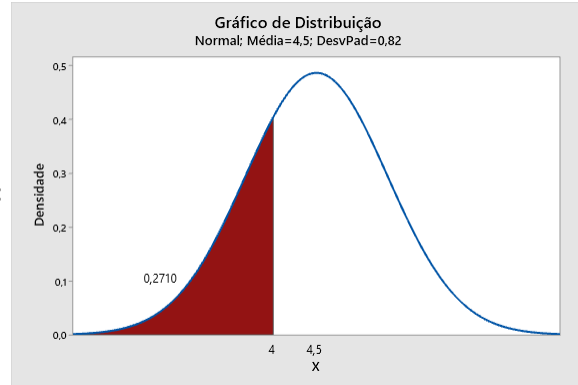
Resultado: **0,9663203**

Exercício 1

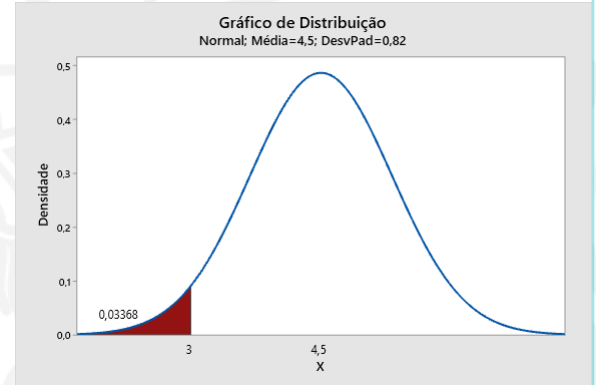
c) Qual é a probabilidade de o porto movimentar entre 3 e 4 milhões de toneladas de carga por semana?



=



-



Exercício 1

c) Qual é a probabilidade de o porto movimentar entre 3 e 4 milhões de toneladas de carga por semana?

$$P(3 < X < 4) = P(X < 4) - P(X < 3) = ?$$

No Excel:

```
=DIST.NORM.N(4;4,5;0,82;VERDADEIRO)
```

```
=DIST.NORM.N(3;4,5;0,82;VERDADEIRO)
```

No R:

```
pnorm(4, 4.5, 0.82, lower.tail = TRUE)
```

```
pnorm(3, 4.5, 0.82, lower.tail = TRUE)
```

No Python:

```
norm.cdf(4, 4.5, 0.82)
```

```
norm.cdf(3, 4.5, 0.82)
```

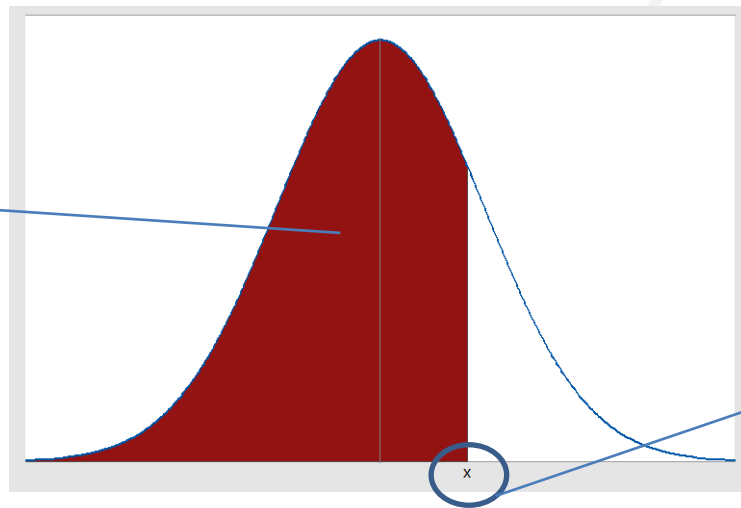
Resultado: **0,2710117 – 0,03367971 = 0,23733199**

Utilizando o Excel – Cálculo inverso

Informa o valor de x a partir de uma probabilidade **acumulada**

=INV.NORM.N(probabilidade;media;desvio padrão)

Forneço a probabilidade acumulada até ele!



Quero encontrar esse valor.

Utilizando o R – Cálculo inverso

Informa o valor de x a partir de uma probabilidade **acumulada**

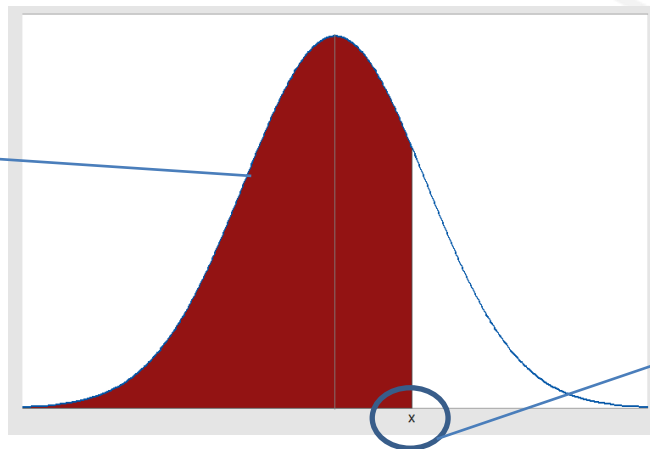
`qnorm(p, m, s, lower.tail = TRUE)`

$p \rightarrow$ representa a probabilidade acumulada até x

$m \rightarrow$ média

$s \rightarrow$ Desvio padrão

Forneço a probabilidade acumulada até ele!



Quero encontrar esse valor.

Utilizando o Python – Cálculo inverso

Informa o valor de x a partir de uma probabilidade **acumulada**

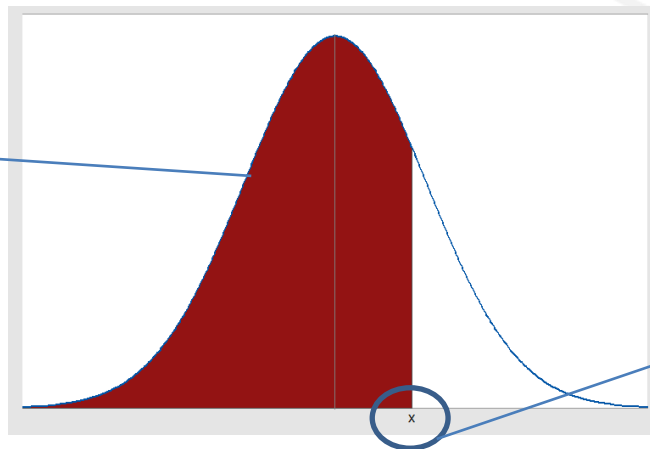
`norm.ppf(p, m, s)`

$p \rightarrow$ representa a probabilidade acumulada até x

$m \rightarrow$ média

$s \rightarrow$ Desvio padrão

Forneço a
probabilidade
acumulada até ele!



Quero encontrar
esse valor.

Exercício 2

Os prazos de duração da gravidez têm distribuição normal com média de 268 dias e desvio padrão de 15 dias. Se definirmos como *prematura* uma criança cujo período de gestação esteja nos 4% inferiores, determine o prazo de gestação que separa as crianças *prematuras* e *não prematuras*.



Exercício 2

X: Tempo de duração da gravidez (em dias)



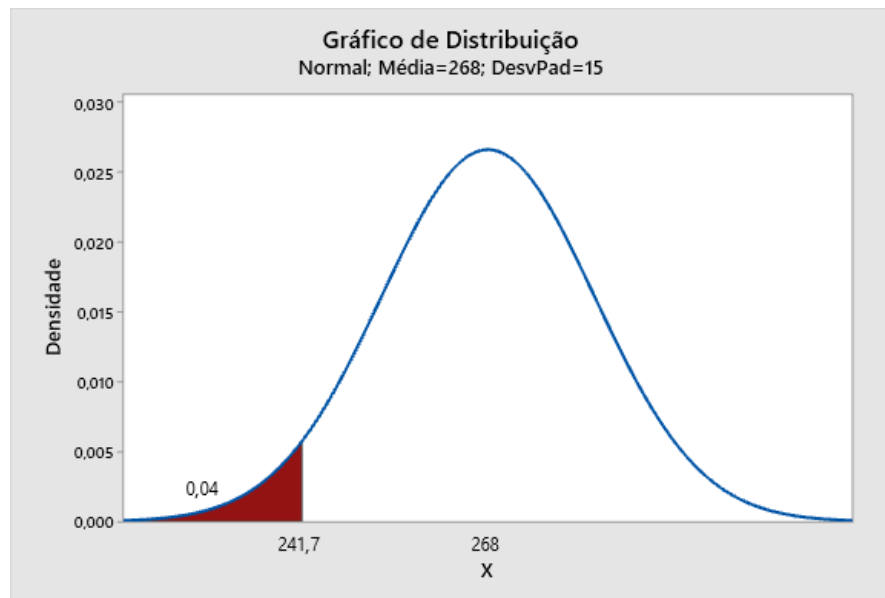
Distribuição Normal



Parâmetros: $\mu = 268$ e $\sigma = 15$.



Exercício 2



$$P(X < t) = 0,04 \rightarrow t = ?$$

No Excel:

=INV.NORM.N(0,04;268;15)

No R:

qnorm(0.04, 268, 15, lower.tail = TRUE)

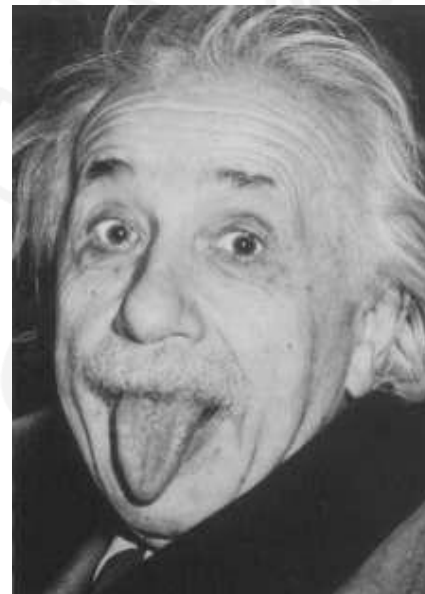
No Python:

norm.ppf(0.04, 268, 15)

Resultado: **241,7397 \approx 242 dias**

Exercício 3

Os QI's têm distribuição normal com média 100 e desvio padrão 15. Definindo como *gênio* uma pessoa no 1% superior dos valores de QI, determine o valor que separa os gênios das pessoas comuns.



Exercício 3

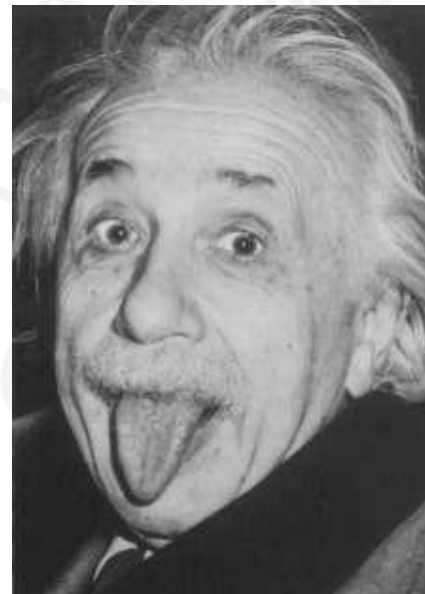
X: Quociente de inteligência (QI)



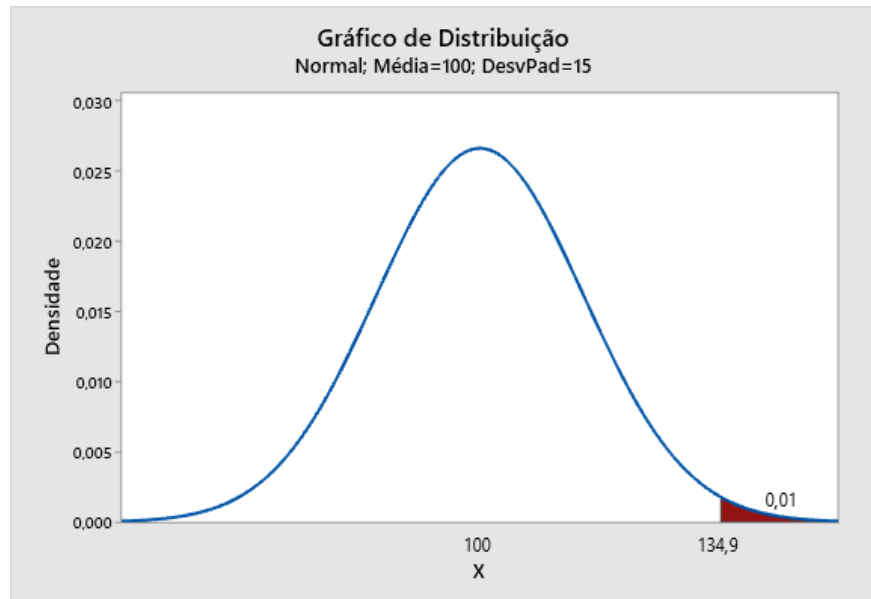
Distribuição Normal



Parâmetros: $\mu = 100$ e $\sigma = 15$.



Exercício 3



$P(X > q) = 0,01 \rightarrow q = ?$
É equivalente a $P(X < q) = 0,99$

No Excel:

`=INV.NORM.N(0,99;100;15)`

No R:

`qnorm(0.99, 100, 15, lower.tail = TRUE)`

No Python:

`norm.ppf(0.99, 100, 15)`

Resultado: **`134,8952`** \approx **`135`**



PUC Minas
Virtual