# ACH2053
# INTRODUÇÃO À ESTATÍSTICA

# 1º Sem/2018

# Breve introdução à distribuição normal

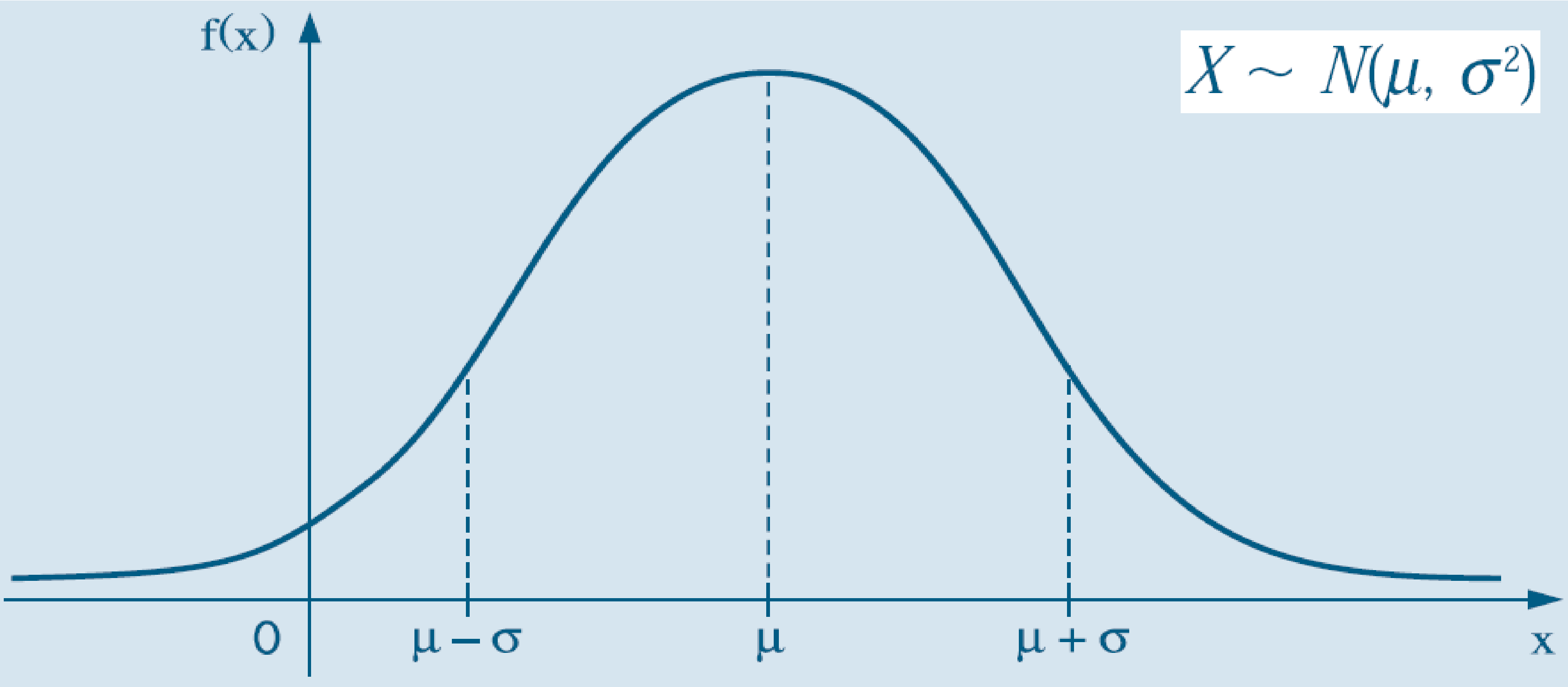
Prof. Marcelo S. Lauretto
marcelolauretto@usp.br
www.each.usp.br/lauretto

Referência:
W.O.Bussab, P.A.Morettin. Estatística Básica, 6ª Edição. São Paulo: Saraiva, 2010 – Capítulo 10

# Distribuição normal

- Uma variável *X* com distribuição normal com média $\mu$ e variância $\sigma^2$ possui função de densidade de probabilidade:

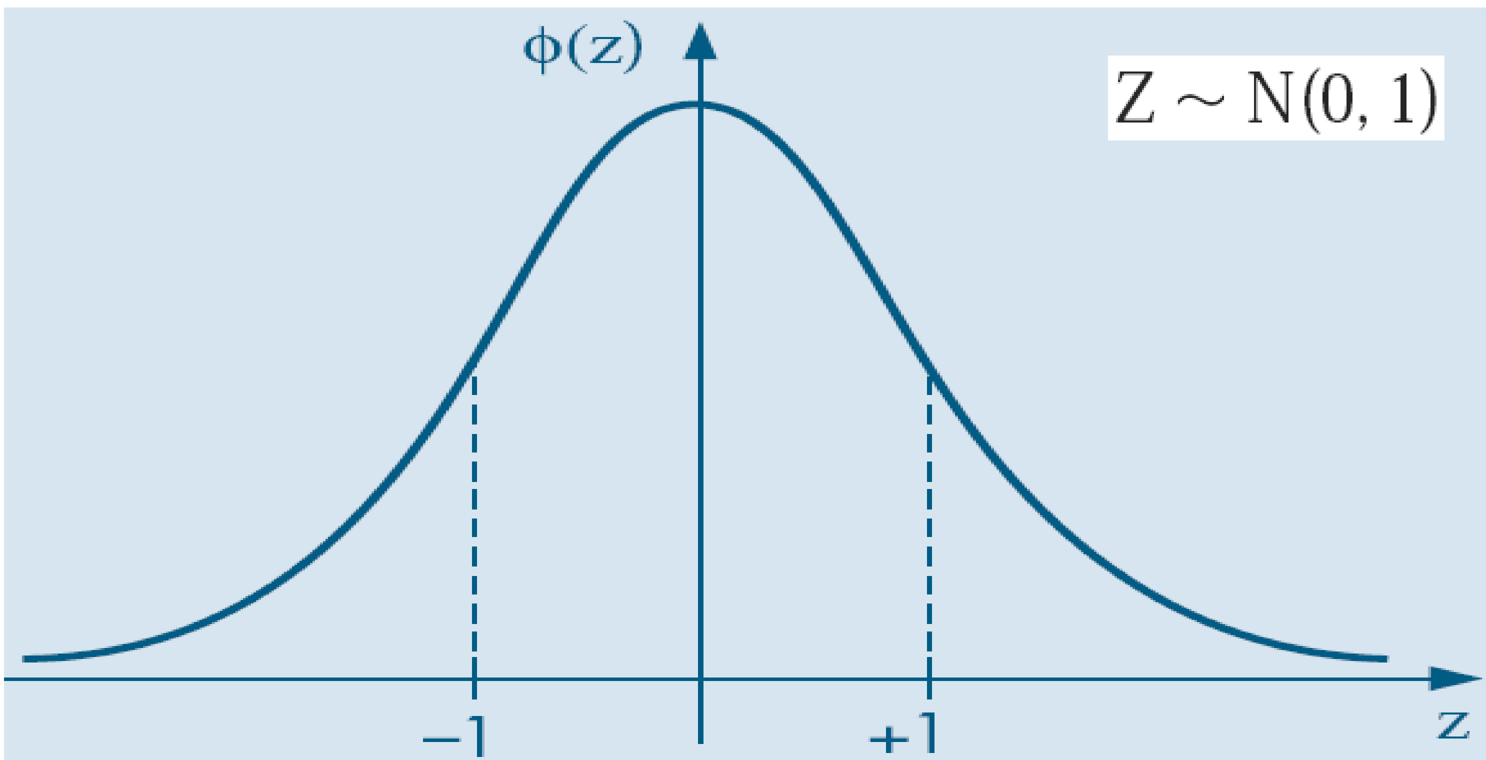
$$f(x;\, \mu,\, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(x-\mu)^2/2\sigma^2},\ -\infty < x < \infty.$$



# Distribuição normal padrão

- Quando $\mu = 0$ e $\sigma^2 = 1$, temos uma distribuição normal padrão:

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

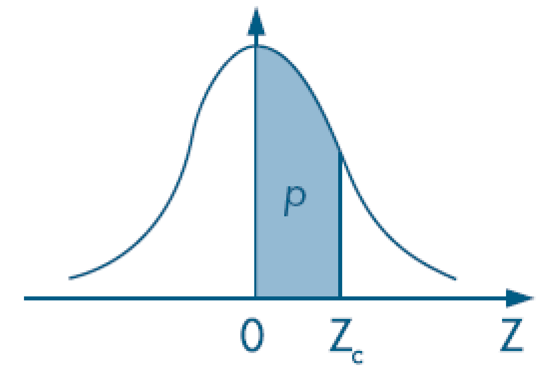


# Tabela da distribuição normal padrão

**Tabela III** — Distribuição Normal Padrão

$Z \sim N(0, 1)$

Corpo da tabela dá a probabilidade $p$, tal que $p = P(0 < Z < Z_c)$




| parte inteira e primeira decimal de $Z_c$ | Segunda decimal de $Z_c$ | | | | | | | | | | parte inteira e primeira decimal de $Z_c$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | |
| | p = 0 | | | | | | | | | | |
| 0,0 | 00000 | 00399 | 00798 | 01197 | 01595 | 01994 | 02392 | 02790 | 03188 | 03586 | 0,0 |
| 0,1 | 03983 | 04380 | 04776 | 05172 | 05567 | 05962 | 06356 | 06749 | 07142 | 07535 | 0,1 |
| 0,2 | 07926 | 08317 | 08706 | 09095 | 09483 | 09871 | 10257 | 10642 | 11026 | 11409 | 0,2 |
| 0,3 | 11791 | 12172 | 12552 | 12930 | 13307 | 13683 | 14058 | 14431 | 14803 | 15173 | 0,3 |
| 0,4 | 15542 | 15910 | 16276 | 16640 | 17003 | 17364 | 17724 | 18082 | 18439 | 18793 | 0,4 |
| 0,5 | 19146 | 19497 | 19847 | 20194 | 20540 | 20884 | 21226 | 21566 | 21904 | 22240 | 0,5 |
| 0,6 | 22575 | 22907 | 23237 | 23565 | 23891 | 24215 | 24537 | 24857 | 25175 | 25490 | 0,6 |
| 0,7 | 25804 | 26115 | 26424 | 26730 | 27035 | 27337 | 27637 | 27935 | 28230 | 28524 | 0,7 |
| 0,8 | 28814 | 29103 | 29389 | 29673 | 29955 | 30234 | 30511 | 30785 | 31057 | 31327 | 0,8 |
| 0,9 | 31594 | 31859 | 32121 | 32381 | 32639 | 32894 | 33147 | 33398 | 33646 | 33891 | 0,9 |
| 1,0 | 34134 | 34375 | 34614 | 34850 | 35083 | 35314 | 35543 | 35769 | 35993 | 36214 | 1,0 |

# Tabela da distribuição normal padrão

Tabela III (cont)

| parte inteira e primeira decimal de $Z_c$ | Segunda decimal de $Z_c$ | | | | | | | | | | parte inteira e primeira decimal de $Z_c$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | |
| 1,1 | 36433 | 36650 | 36864 | 37076 | 37286 | 37493 | 37698 | 37900 | 38100 | 38298 | 1,1 |
| 1,2 | 38493 | 38686 | 38877 | 39065 | 39251 | 39435 | 39617 | 39796 | 39973 | 40147 | 1,2 |
| 1,3 | 40320 | 40490 | 40658 | 40824 | 40988 | 41149 | 41309 | 41466 | 41621 | 41774 | 1,3 |
| 1,4 | 41924 | 42073 | 42220 | 42364 | 42507 | 42647 | 42786 | 42922 | 43056 | 43189 | 1,4 |
| 1,5 | 43319 | 43448 | 43574 | 43699 | 43822 | 43943 | 44062 | 44179 | 44295 | 44408 | 1,5 |
| 1,6 | 44520 | 44630 | 44738 | 44845 | 44950 | 45053 | 45154 | 45254 | 45352 | 45449 | 1,6 |
| 1,7 | 45543 | 45637 | 45728 | 45818 | 45907 | 45994 | 46080 | 46164 | 46246 | 46327 | 1,7 |
| 1,8 | 46407 | 46485 | 46562 | 46638 | 46712 | 46784 | 46856 | 46926 | 46995 | 47062 | 1,8 |
| 1,9 | 47128 | 47193 | 47257 | 47320 | 47381 | 47441 | 47500 | 47558 | 47615 | 47670 | 1,9 |
| 2,0 | 47725 | 47778 | 47831 | 47882 | 47932 | 47982 | 48030 | 48077 | 48124 | 48169 | 2,0 |
| 2,1 | 48214 | 48257 | 48300 | 48341 | 48382 | 48422 | 48461 | 48500 | 48537 | 48574 | 2,1 |
| 2,2 | 48610 | 48645 | 48679 | 48713 | 48745 | 48778 | 48809 | 48840 | 48870 | 48899 | 2,2 |
| 2,3 | 48928 | 48956 | 48983 | 49010 | 49036 | 49061 | 49086 | 49111 | 49134 | 49158 | 2,3 |
| 2,4 | 49180 | 49202 | 49224 | 49245 | 49266 | 49286 | 49305 | 49324 | 49343 | 49361 | 2,4 |

# Padronização

- Se Quando $X \sim N(\mu, \sigma^2)$, então a variável aleatória definida por

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

terá uma distribuição normal padrão

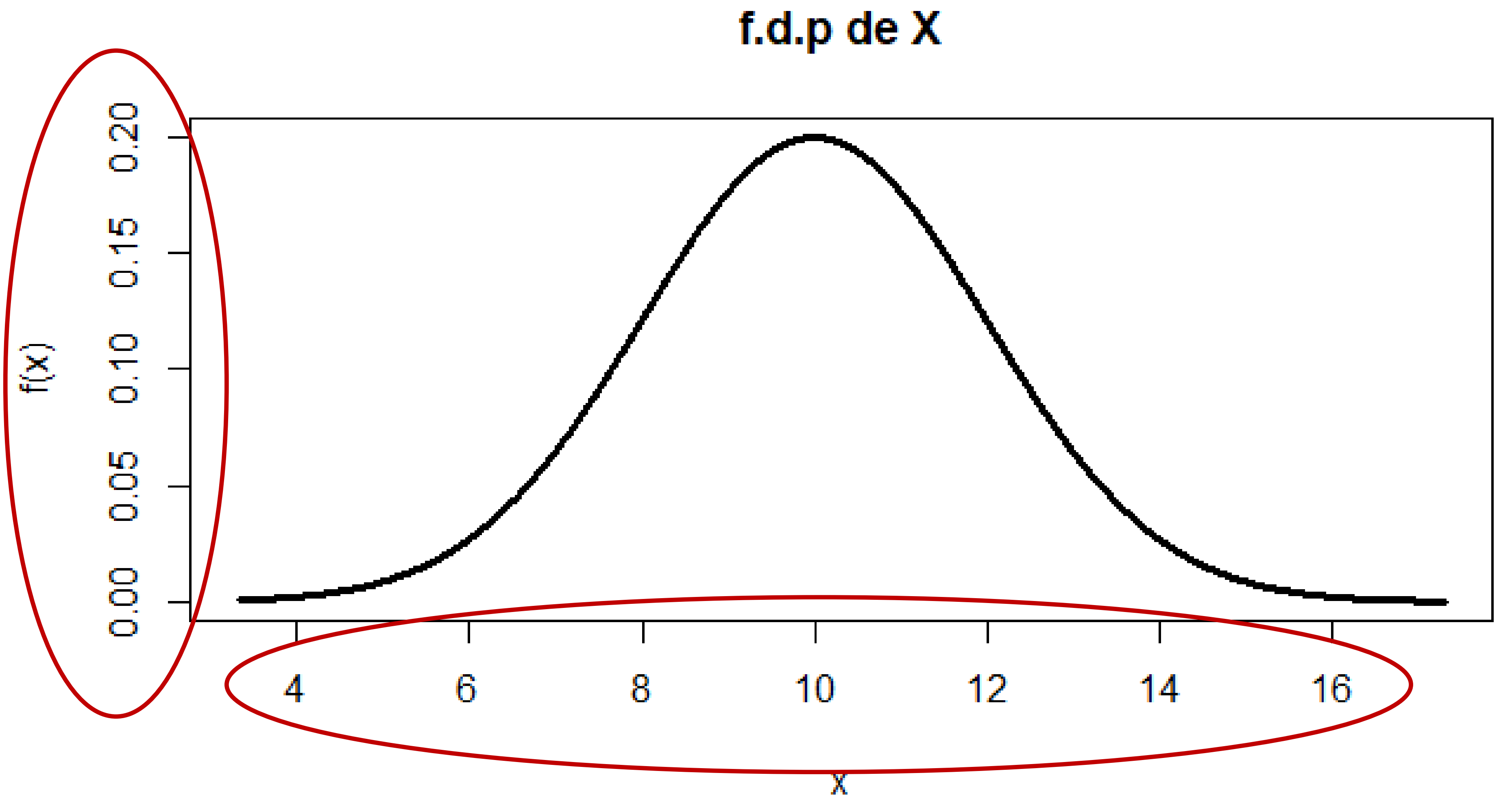
- Por quê?
  - $E[Z] = E\left[\frac{X-\mu}{\sigma}\right] = \frac{E[X]-\mu}{\sigma} = \frac{\mu-\mu}{\sigma} = 0$
  - $\mathrm{Var}[Z] = \mathrm{Var}\left[\frac{X-\mu}{\sigma}\right] = \frac{1}{\sigma^2}(\mathrm{Var}[X] - \mathrm{Var}\,[\mu]) = \frac{1}{\sigma^2}(\sigma^2 - 0) = 1$
  - Importante: Note que

$$\Pr(x_{min} < X < x_{max}) = \Pr\left(\frac{x_{min} - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{x_{max} - \mu}{\sigma}\right)$$

$$= \Pr\left(\frac{x_{min} - \mu}{\sigma} < Z < \frac{x_{max} - \mu}{\sigma}\right)$$

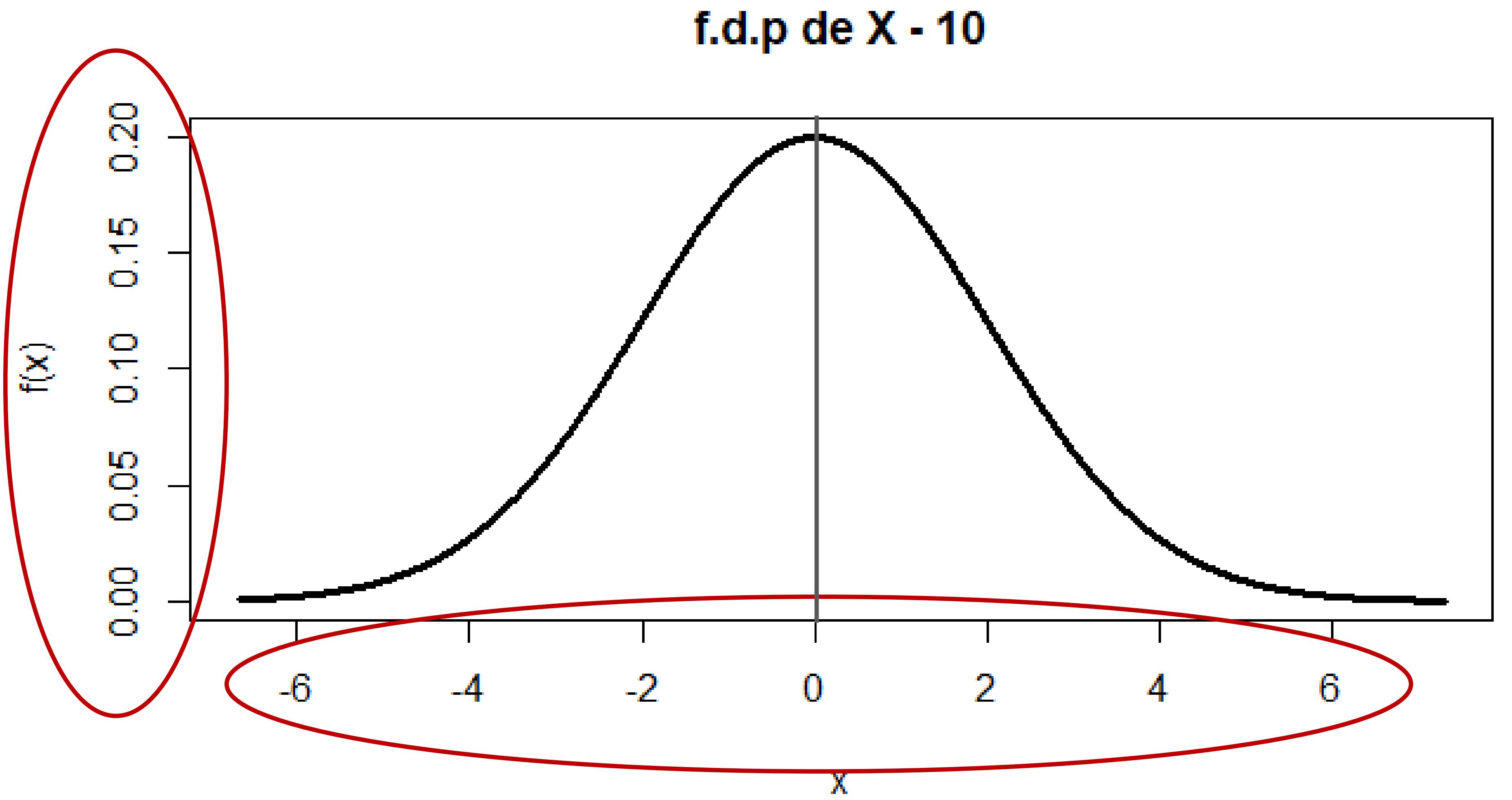
# Padronização

- Exemplo: $X \sim N(10, 4) \longrightarrow Z = \frac{X-10}{2}$



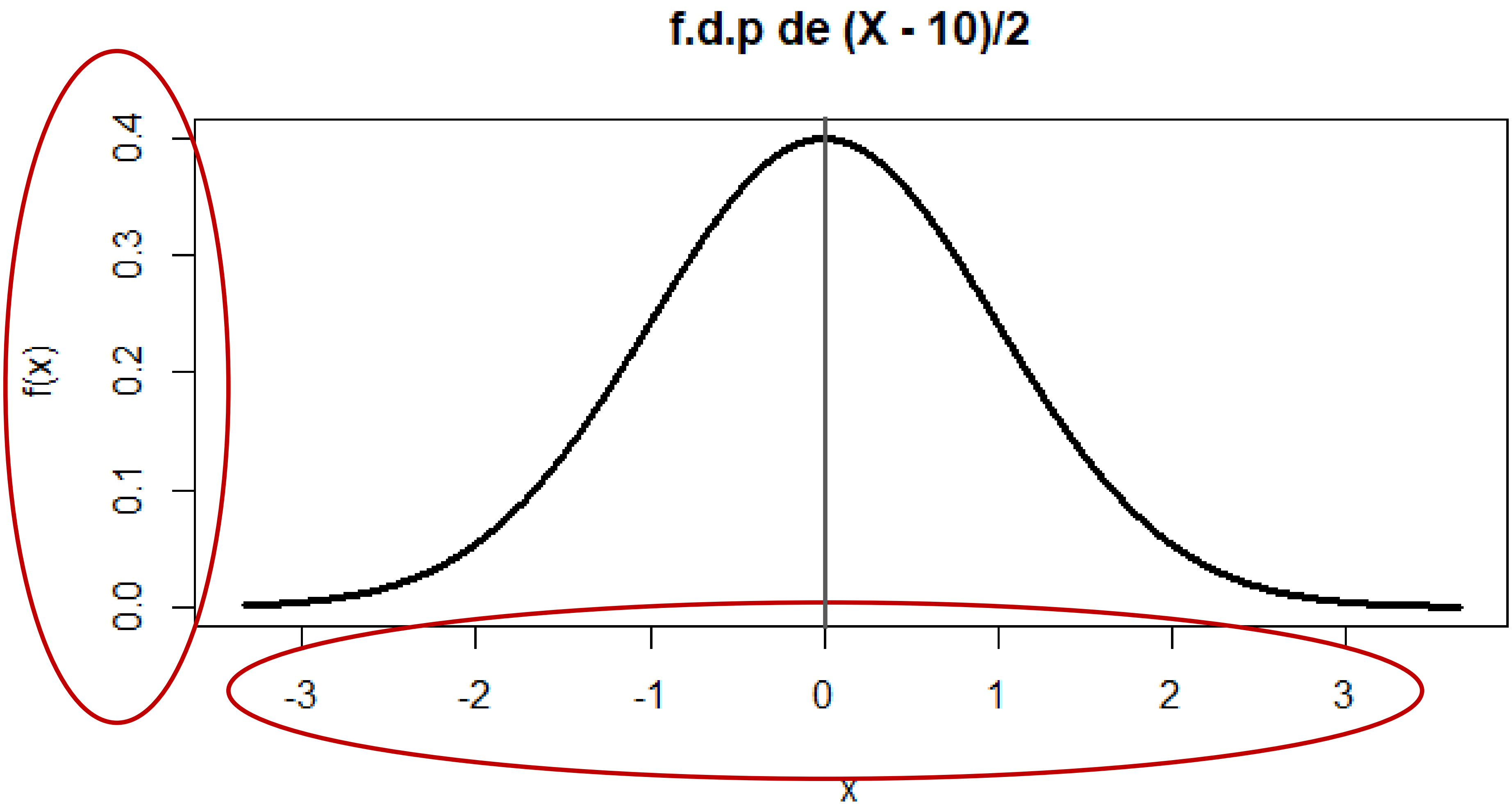
# Padronização

- Exemplo: $X \sim N(10, 4) \longrightarrow Z = \frac{X-10}{2}$



# Padronização

- Exemplo: $X \sim N(10,4) \longrightarrow Z = \frac{X-10}{2}$



# Padronização

- A transformação *Z* é usualmente denominada *padronização* de *X*
  - Interpretação: Z é a distância de X em relação à média $\mu$, medida em desvios-padrão.
    - Por exemplo, se uma observação $x = 6$ é obtida a partir de uma distribuição normal padrão com média $\mu = 10$ e desvio padrão $\sigma = 2$, então $z = \frac{x-\mu}{\sigma} = \frac{6-10}{2} = -2$ indica que $x$ está a dois desvios-padrão de distância de $\mu$ (o que ocorreria também se $x = 14$)
  - Facilita cálculos manuais, pois pode-se usar a tabela da distribuição normal padrão para calcular as probabilidades de interesse
  - Desnecessária se os cálculos forem realizados em ambientes como Excel, R etc, que implementam a distribuição Normal geral ($\mu$ e $\sigma$ quaisquer)

# Exemplo

- Exemplo: Grear Tire Company
  - A fabricante Grear Tire Company desenvolveu um novo pneu radial que será vendido em uma cadeia de lojas de desconto
  - Por ser um novo produto, a companhia acredita que a garantia de quilometragem oferecida com o produto será um fator importante de aceitação
    - A cláusula de garantia prevê um desconto na troca caso um pneu vendido dure menos do que uma certa quilometragem mínima garantida
  - Para planejar a política de garantia, deseja-se obter informação sobre a quilometragem que o pneu deve rodar
  - Testes preliminares com uma amostra do novo pneu indicaram que a duração *X* (em Km) tem uma média $\mu = 36.500$ Km e um desvio padrão $\sigma = 5.000$ Km
  - Adicionalmente, os dados coletados indicaram que a distribuição normal é uma assunção razoável

# Exemplo

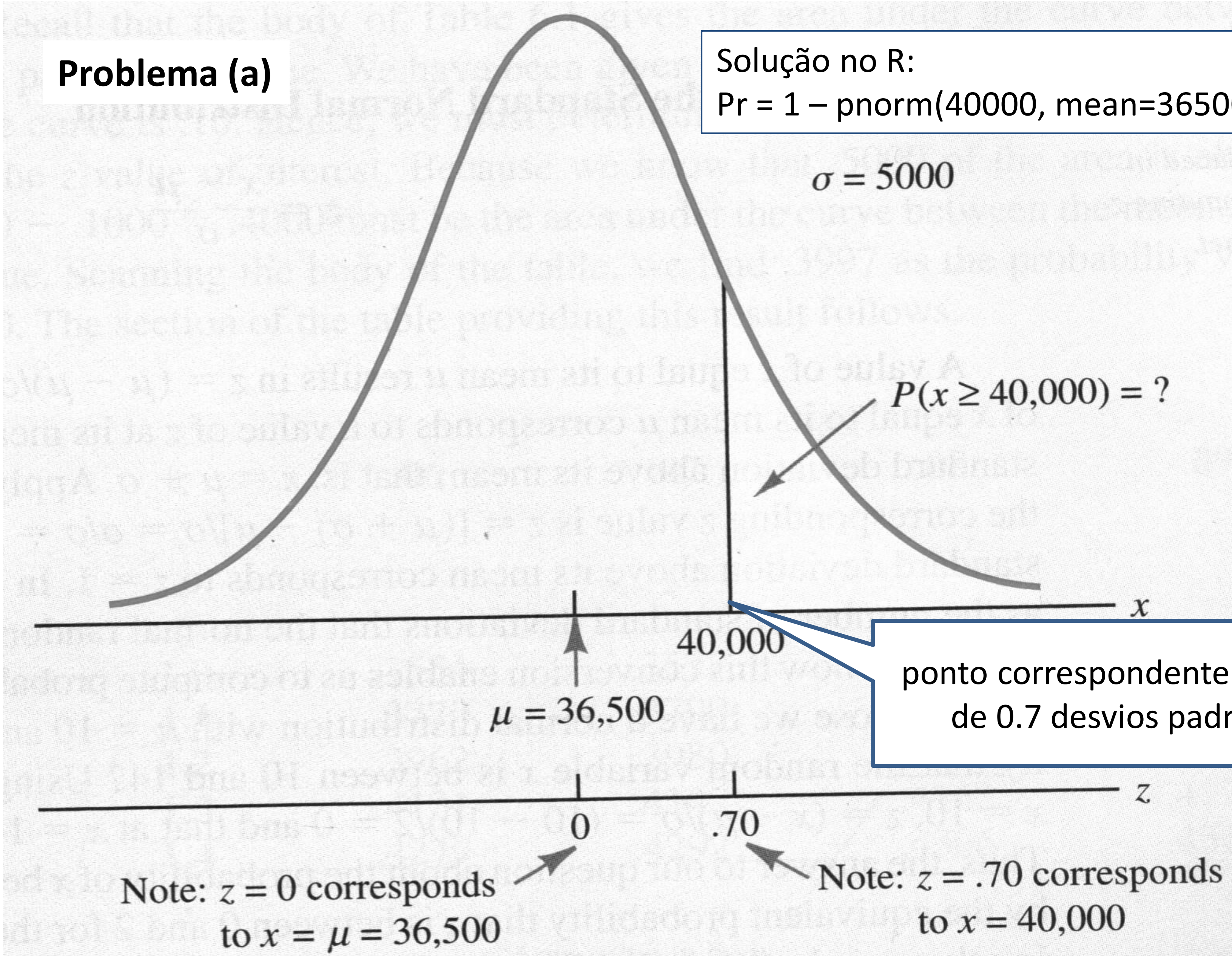
- Exemplo: Grear Tire Company
  a) Considerando que uma das peças publicitárias do novo pneu pretende anunciar que o produto poderá durar 40 mil km (o que seria um bom apelo de vendas), qual a probabilidade de um pneu vendido chegar de fato a essa quilometragem?
  b) Qual deve ser o valor de quilometragem garantida de modo que no máximo 10% das unidades vendidas sejam elegíveis ao desconto?

- A questão (a) pode ser respondida analisando a distribuição de distribuição acumulada, $F$
- A questão (b) pode ser respondida através da função quantil (ou função inversa) $F^{-1}$ da distribuição normal

# Exemplo

**Problema (a)**

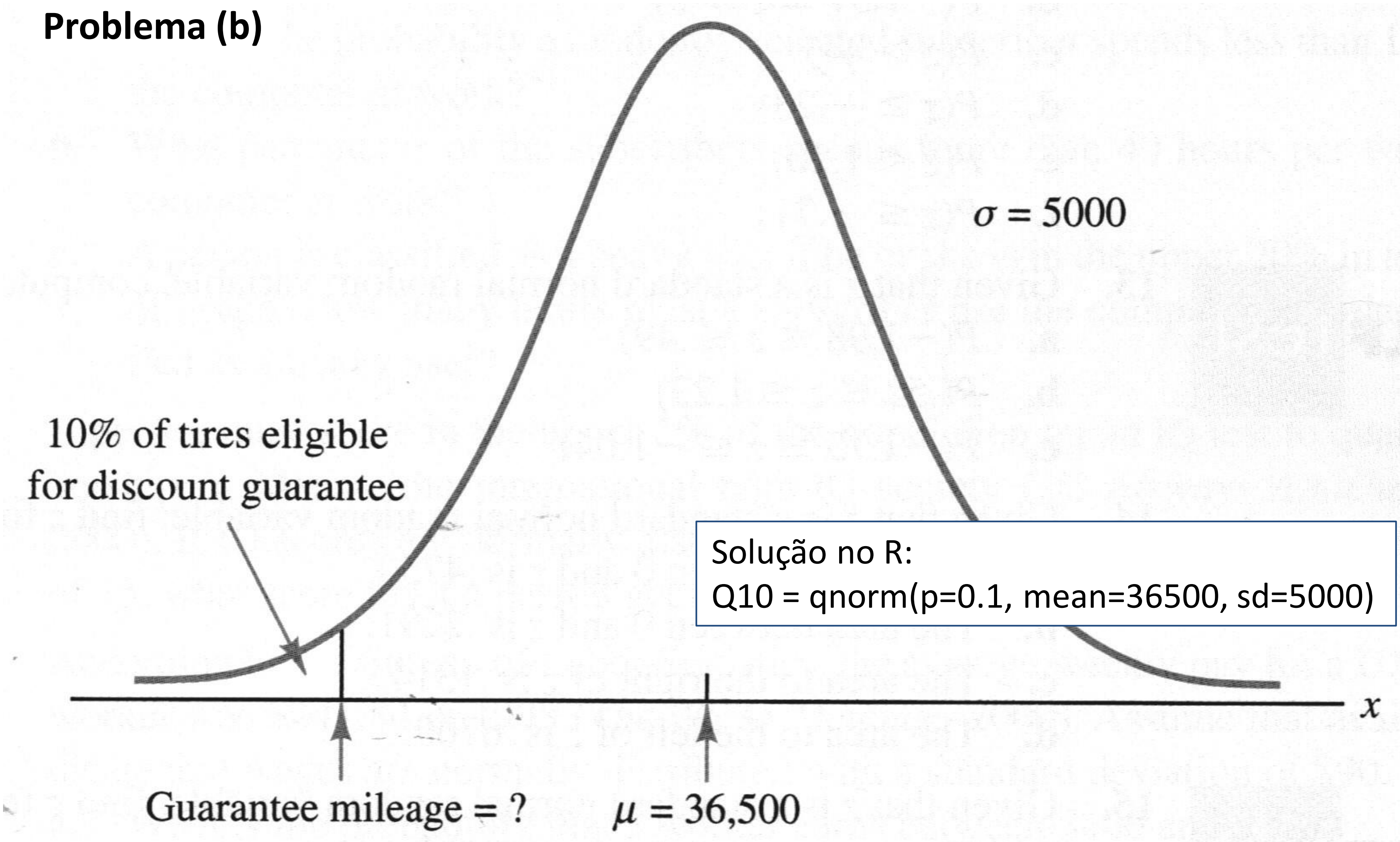
Solução no R:
Pr = 1 – pnorm(40000, mean=36500, sd=5000)




ponto correspondente à distância de 0.7 desvios padrão de $\mu$

# Exemplo

**Problema (b)**




Solução no R:
Q10 = qnorm(p=0.1, mean=36500, sd=5000)

# Exemplo

- Exemplo: Grear Tire Company

  a) Considerando que uma das peças publicitárias do novo pneu pretende anunciar que o produto poderá durar 40 mil km (o que seria um bom apelo de vendas), qual a probabilidade de um pneu vendido chegar de fato a essa quilometragem?

  - Queremos calcular

    $$\Pr(X \geq 40000) = 1 - \Pr(X < 40000) = 1 - F(40000)$$

  - No Excel, basta usar a fórmula

    =1 – DIST.NORM.N(40000,36500,5000,VERDADEIRO)

    o que resulta em aproximadamente 0.242 = 24.2%

  - No R: 1 – pnorm(40000,mean=36500, sd=5000)

  - Para o cálculo manual, devemos padronizar X,

    $$Z = \frac{X - \mu}{\sigma} = \frac{40000 - 36500}{5000} = 0.7$$

    Pela Tabela III, sabemos que $\Pr(0 < z < 0.7) = 0.258$ e portanto $\mathrm{F}(0.7) = \Pr(-\infty < z < 0.7) = 0.758$; como estamos interessados na área à direita de 0.7, devemos subtrair $1 - 0.758 = 0.242 = 24.2\%$

# Exemplo

- Exemplo: Grear Tire Company

  b) Qual deve ser o valor de quilometragem garantida de modo que no máximo 10% das unidades vendidas sejam elegíveis ao desconto?

  - Nosso interesse é calcular $F^{-1}(p)$
  - Em Excel, basta usar a fórmula

    =INV.NORM.N(0.1, 36500, 5000)

    o que resulta em 30092 km
  - No R: qnorm(p=0.1, mean=36500, sd=5000)
  - Para o cálculo manual:
    - devemos observar que 40% da área deve estar entre a média e o valor da quilometragem de garantia;
    - Procuramos pelo valor 0.4 na Tabela III e descobrimos que esta área está a aproximadamente 1.28 desvios padrões abaixo da média, ou seja, $z = -1.28$;

# Exemplo

- Exemplo: Grear Tire Company

  b) Qual deve ser o valor de quilometragem garantida de modo que no máximo 10% das unidades vendidas sejam elegíveis ao desconto?

  - Para o cálculo manual (cont):
    - Para encontrar a quilometragem correspondente a $z = -1.28$, temos:

      $$\mathrm{z} = \frac{x - \mu}{\sigma} = -1.28 \quad \Longrightarrow x - \mu = -1.28\sigma \Longrightarrow x = \mu - 1.28\sigma$$

    - Com $\mu = 36.500$ e $\sigma = 5.000$, temos

      $x = \mu - 1.28\sigma = 36500 - 1.28\ (5000)$ =30100

  - Assim, uma garantia de 30.100 km satisfaria o requisito de aproximadamente 10% dos pneus serem elegíveis para o desconto
  - Com base nessa informação, seria razoável que a empresa estabeleça uma quilometragem de garantia de 30.000 km.