Variável aleatória contínua

Gilberto Pereira Sassi

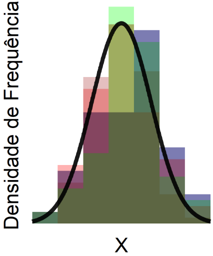
Universidade Federal da Bahia
Instituto de Matemática e Estatística
Departamento de Estatística

Objetivo

Caracterização probabilística de variáveis quantitativas contínuas. Neste caso, não existe interesse interesse em atribuir probabilidades a cada valor individual da variável e estamos interessados em atribuir probabilidades para intervalos da variável.

Para uma variável quantitativa contínua X, podemos construir o histograma que é válido para uma amostra específica. Para uma outra amostra da variável quantitativa contínua X na mesma população, podemos obter um histograma ligeiramente diferente. A ideia é obter uma curva ou uma função que aproxime bem o formato de todos histogramas, conforme ilustrado na figura. Chamamos esta curva de função densidade de probabilidade ou simplesmente função densidade.

Figura 1: Histogramas sobrepostos para diversas amostras da variável X. A linha azul é a função densidade de probabilidade da variável quantitativa contínua X.



Definição

Distribuição de probabilidade

Uma distribuição de probabilidade para uma variável quantitativa contínua X é estabelecido quando definimos:

1. os valores possíveis para a variável;
2. a função densidade de probabilidade.

Quando temos uma função densidade de probabilidade associada a uma variável X, dizemos que X é uma variável aleatória contínua para deixar claro que temos um modelo de probabilidade para X.

Definição

Dizemos que $f(x)$ é uma função densidade de probabilidade para um variável aleatória contínua X se satisfaz duas condições:

1. $f(x) \geq 0$;
2. A área delimitada por $f(x)$ e o eixo x é igual a 1.

Observação

1) Note que $P(X = x) = 0$, para $x \in [a, b]$, logo $P(a < X \leq b) = P(a < X < b) = P(a \leq X < b) = P(a \leq X \leq b)$.
2) A área sob a curva da função densidade de probabilidade é igual a $P(a < X < b)$ como ilustrado na figura 2.

Área sob a curva como probabilidade.

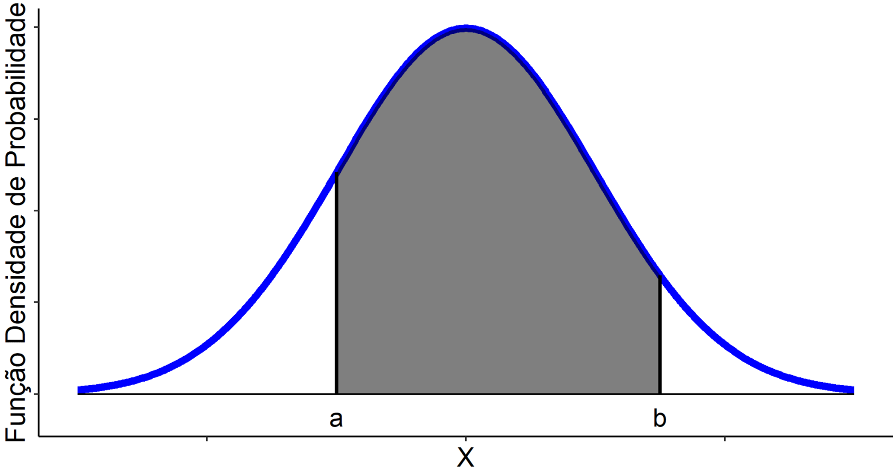


Figura 2: Probabilidade de uma variável aleatória contínua X estar entre a e b. A área pintada no gráfico é o valor de $P(a < X < b)$.

Exemplo

Num teste tradicional com crianças, o tempo em minutos para a realização de uma bateria de questões de raciocínio verbal e lógico é uma variável aleatória contínua T com função densidade de probabilidade dada por

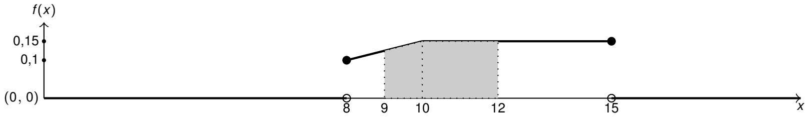
$$f(t) = \begin{cases} \dfrac{t-4}{40}, & \text{se } 8 \leq t < 10, \\ \dfrac{3}{20}, & \text{se } 10 \leq t \leq 15, \\ 0, & \text{caso contrário.} \end{cases}$$

Qual a probabilidade de uma criança responder a bateria de teste entre 9 e 12 minutos?

Exemplo

$P(9 < X \leq 12)$ é área delimitada pelo gráfico conforme figura 3.

Figura 3: Função denside.



$$P(9 < X \leq 12) = 1 \cdot \frac{0,15 + 0,1}{2} + 2 \cdot 0,15$$
$$= 0,4375.$$

Definição

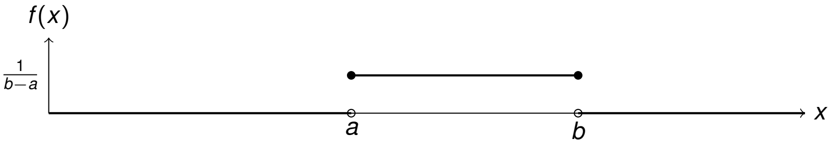
- Média: $\mathrm{E}(X) = \mu = \int_{-\infty}^{\infty} xf(x)dx$;
- Mediana: valor Md com a propriedade $P(X \geq Md) \geq 0,5$ e $P(X \leq Md) \geq 0,5$;
- Quantil de ordem p: valor $Q(p)$ com a propriedade $P(X \leq Q(p)) \geq p$ e $P(X \geq Q(p)) \geq 1 - p$;
- Moda: ponto de máximo de $f(x)$, ou seja, valor Mo tal que $f(x) \leq f(Mo), \forall x$;
- Variância: $\mathrm{Var}(X) = \sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx$.

Distribuição uniforme contínua

Uma variável aleatória contínua X tem distribuição uniforme contínua no intervalo $[a, b]$ se sua função densidade de probabilidade é dada por:

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & \text{se } a \leq x \leq b, \\ 0, & \text{caso contrário .} \end{cases}$$

Figura 4



Notação: $X \sim U[a, b]$.

Propriedades – distribuição uniforme contínua

i. **Média** $\mu = \mathrm{E}(x) = \dfrac{b+a}{2}$;

ii. **Variância:** $\sigma^2 = \mathrm{Var}(X) = \dfrac{(b-a)^2}{12}$;

iii. **Mediana:** $Md = \dfrac{b+a}{2}$;

iv. **Quantil:** $Q(p) = p \cdot (b-a) + a$;

v. **Moda:** Mo : é qualquer número em $[a, b]$;

vi. **FDA:** $F(x) = \begin{cases} 0, & \text{se } x < a, \\ \frac{x-a}{b-a}, & \text{se } a \leq x \leq b \\ 1, & \text{se } x \geq b \end{cases}$

Exemplo

Admite-se que uma pane elétrica pode ocorrer em qualquer ponto de uma rede elétrica de 10km.

a) Qual a probabilidade da pane ocontecer nos primeiros 500 metros;

b) O custo do reparo da rede depende da distância do centro do serviço ao local da pane. Considere que o centro de serviço está na origem da rede e que o custo é dado pela tabela

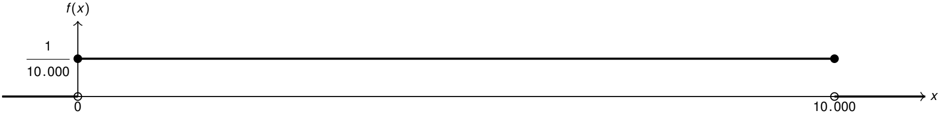
Distância em km	Custo
$[0, 3)$	$R\$200,00$
$[3, 8)$	$R\$400,00$
$[8, 10]$	$R\$1000,00$

Qual o custo médio para reparar a rede?

Solução:

A função de probabilidade é dada por

Figura 5



a) $P(0 \leq X \leq 500) = \dfrac{500}{10.000} = \dfrac{1}{20} = 0,05;$

b)
- $P(0 \leq X \leq 3.000) = \dfrac{3.000}{10.000} = 0,3;$
- $P(3.000 \leq X \leq 8.000) = \dfrac{5.000}{10.000} = \dfrac{1}{2} = 0,5$
- $P(8.000 \leq X) = \dfrac{2.000}{10.000} = \dfrac{1}{5} = 0,2$

Exemplo - continuação

Ou seja, temos uma tabela de distância, custo e probabilidade dada por

Distância em km	Custo	Probabilidade de pane dentro da distância
$[0,3)$	$R\$200,00$	$0,3$
$[3,8)$	$R\$400,00$	$0,5$
$[8,10]$	$R\$1000,00$	$0,2$

Então podemos estabelecer uma variável aleatória discreta Y, custo de reparo, com valores possíveis 200, 400 e 1000 e função de probabilidade

$$\begin{aligned} f(200) &= 0,3; \\ f(400) &= 0,5; \\ f(1000) &= 0,2. \end{aligned}$$

e o custo médio é dado por

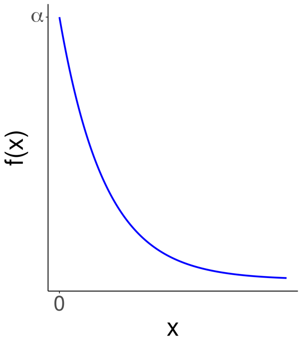
$$\mathrm{E}(Y) = 0,3 \cdot 200 + 0,5 \cdot 400 + 0,2 \cdot 1000 = 460.$$

Ou seja, o custo médio de reparo para esta rede elétrica é de $R\$460,00$.

Distribuição exponencial

Uma variável aleatória contínua X segue o modelo exponencial com parâmetro α se sua densidade é dada por $f(x) = \alpha e^{-\alpha \cdot x}$, para $x \geq 0$.

Figura 6



Notação: $X \sim Exp(\alpha)$

Propriedades – distribuição exponencial

i. **Média** $\mathrm{E}(X) = \mu = \dfrac{1}{\alpha}$;

ii. **Variância:** $\mathrm{Var}(X) = \sigma^2 = \dfrac{1}{\alpha^2}$;

iii. **Mediana:** $Md(X) = \dfrac{\log 2}{\alpha}$;

iv. **Quantil:** $Q(p) = \dfrac{-\ln(1-p)}{\alpha}$;

v. **Moda:** $Mo = 0$;

vi. **FDA:** $F(x) = \begin{cases} 0, & \text{se } x < 0, \\ 1 - \exp(-\alpha \cdot x), & \text{se } X \geq 0. \end{cases}$

Distribuição exponencial

Uso

Modelagem de tempo até ocorrer um exemplo. Por exemplo: tempo até o óbito, tempo até uma falha de um equipamento, tempo até solicitação ou ligação, etc.

Cálculo da área sob a curva

Seja $X \sim Exp(\alpha)$, seja a e b com $a < b < \infty$ então

- Se $a > 0$, então $\int_a^b \alpha e^{-\alpha \cdot x} dx = e^{-\alpha \cdot a} - e^{-\alpha \cdot b}$;
- Se $a \leq 0$, então $\int_a^b \alpha e^{-\alpha \cdot x} dx = 1 - e^{-\alpha \cdot b}$.

Exemplo

Sabe-se que um paciente em estágio avançado de uma certa enfermidade vive em média apenas em 120 dias. Qual a probabilidade de um paciente morrer antes de 90 dias?

Solução: Pelo enunciado do exercício, temos que $\mathrm{E}(X) = \frac{1}{\alpha} = 120$, ou seja, $\alpha = \frac{1}{120} = 0,008$. Então,

$$P(X \leq 90) = 1 - \exp(-0,008 \cdot 90) = 0,53.$$

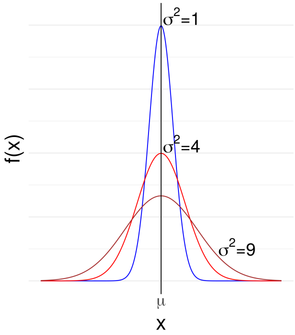
ou seja, probabilidade do paciente morrer antes de 90 dias é 53%.

Distribuição normal

Dizemos que uma variável aleatória contínua X tem distribuição normal com parâmetro μ e σ^2 se sua função densidade de probabilidade é dada por:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad x \in \mathbb{R}.$$

Figura 7



Notação: $X \sim N(\mu, \sigma^2)$.

Propriedade: distribuição normal

i) $f(x)$ é simétrica em relação a μ;

ii) $f(x)$ tende a 0 quando $x \to \pm\infty$;

iii) $\mathrm{E}(X) = \mu$;

iv) $\mathrm{Var}(X) = \sigma^2$;

v) A moda e a mediana de X é μ;

vi) $P(a < X \leq b) = P(a < X < b) = P(a \leq X < b) = P(a \leq X \leq b)$;

vii) Se $-\infty \leq a < b \leq \infty$, então

   a) $P(a < X < b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$;

   b) $P(X < b) = \Phi\left(\frac{b-\mu}{\sigma}\right)$;

   em que os valores de $\Phi(z)$ são tabelados;

viii) Se $c < 0$, então $\Phi(c) = 1 - \Phi(-c)$;

ix) Se $c = -\infty$, então $\Phi(c) = 0$;

x) Se $c = \infty$, então $\Phi(c) = 1$.

Exemplo

Doentes, sofrendo de certa moléstia, são submetidos a um tratamento intensivo cujo tempo de cura foi modelado por uma densidade Normal de média 15 e desvio padrão 2. Qual a proporção desses pacientes demoram mais de 17 dias para se recuperar? Qual o tempo máximo para a recuperação de 25% dos pacientes?

Solução: Note que $X \sim N(15, 4)$.

- $$\begin{aligned} P(17 > X) &= 1 - P(X \leq 17) \\ &= 1 - \Phi\left(\frac{17 - 15}{\sqrt{4}}\right) \\ &= \Phi(-1) = 0,1587. \end{aligned}$$

  Ou seja, 16% dos porcentagens pacientes demoram mais de 17 dias para se recuperar;
- $P(X \leq t) = \Phi\left(\dfrac{t - 15}{2}\right) = 0,25$, então e $\dfrac{t-15}{2} = -0,68$ e $t = -0,68 \cdot 2 + 15 = 13,64$. Ou seja, o tempo máximo para recuperação de 25% é de 2,651 dias.

Distribuição amostral: motivação

Imagine que um professor tem uma turma com 30 alunos. As notas finais destes 30 alunos estão na Tabela 1. Este professor está com tempo limitado e decidiu analisar o desempenho de 5 alunos ao final do curso. Existem 142.506 maneiras de selecionar esses cinco alunos. Na Tabela 2, mostramos dez amostras diferentes com cinco alunos. Note que cada amostra tem uma média diferente. A ideia é que a média é uma variável (valor diferente em cada amostra) que denotamos por $\bar{X}$.

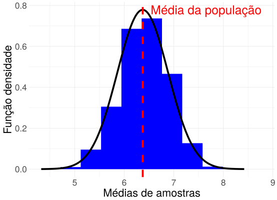
7,29	7,19	7,15	5,54	5,93	5,53	6,44	6,27	8,16	5,72
4,84	4,63	6,11	7,10	3,37	7,36	6,70	5,70	6,31	7,64
5,89	8,82	7,77	7,93	5,24	6,08	5,77	6,57	6,00	6,14

Tabela 1: Turma com 30 alunos.

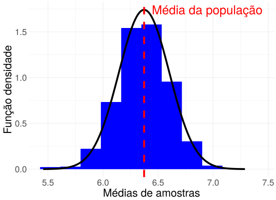
Amostras	Aluno 1	Aluno 2	Aluno 3	Aluno 4	Aluno 5	Média da amostra
Amostra 1	7,19	8,82	5,54	6,70	6,11	6,87
Amostra 2	7,36	4,84	6,31	6,27	7,29	6,41
Amostra 3	5,77	6,57	6,44	8,16	7,93	6,97
Amostra 4	6,44	7,29	5,77	3,37	6,11	5,80
Amostra 5	6,14	7,36	3,37	6,70	7,10	6,13
Amostra 6	8,82	6,31	7,36	5,77	5,72	6,80
Amostra 7	7,10	7,93	4,84	6,44	5,93	6,45
Amostra 8	7,10	5,72	7,36	5,77	4,84	6,16
Amostra 9	6,44	6,14	7,64	6,08	5,70	6,40
Amostra 10	6,00	7,77	5,53	5,24	7,15	6,34

Tabela 2: Dez amostras com cinco alunos com a média.

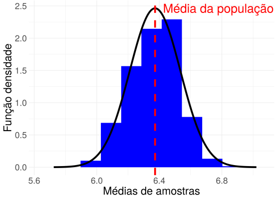
Distribuição amostral de médias de notas.



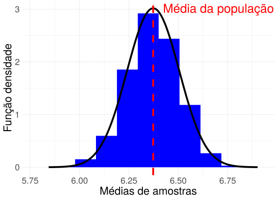
(a) $n = 5$



(b) $n = 25$



(c) $n = 50$



(d) $n = 75$

Figura 8: Médias das amostras: demonstração do teorema central do limite.

Distribuição amostral: motivação

Considere a variável discreta X com suporte e função de probabilidade dada pela Tabela 3. Na Tabela 4, apresentamos dez amostras com cinco valores. Note que cada amostra tem uma média que não precisa ser um número inteiro. A ideia é que a média uma variável aleatória (valor diferente em cada amostra) que denotamos por $\bar{X}$.

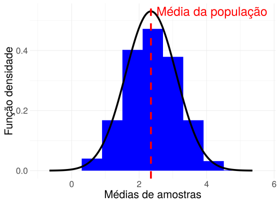
Tabela 3: Variável discreta X com suporte e função de probabilidade.

x	0	1	2	3	4	5
$f(x)$	$0,1$	$0,4$	$0,05$	$0,05$	$0,3$	$0,1$

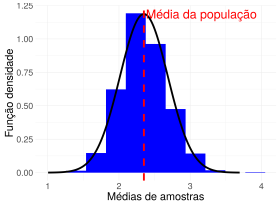
Tabela 4: Dez amostras de uma variável quantitativa.

Amostras	Valor 1	Valor 2	Valor 3	Valor 4	Valor 5	Média
Amostra 1	4	1	1	4	1	2,20
Amostra 2	1	0	1	1	4	1,40
Amostra 3	4	4	5	3	4	4,00
Amostra 4	4	1	5	4	4	3,60
Amostra 5	4	5	1	4	1	3,00
Amostra 6	0	1	0	1	0	0,40
Amostra 7	5	4	4	4	1	3,60
Amostra 8	2	3	5	5	2	3,40
Amostra 9	4	4	3	4	1	3,20
Amostra 10	1	1	4	1	5	2,40

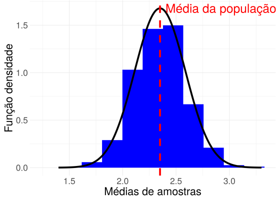
Distribuição amostral de médias da variável discreta X.



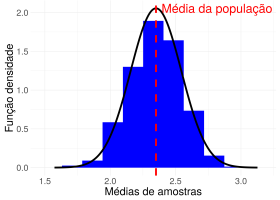
(a) $n = 5$



(b) $n = 25$



(c) $n = 50$



(d) $n = 75$

Figura 9: Médias das amostras: demonstração do teorema central do limite.

Teorema central do limite

Ideia

Para um tamanho de amostra suficientemente grande, a distribuição de $\bar{X}$ pode ser aproximada por uma distribuição normal, independente do modelo de probabilidade de X_i.

Teorema central do limite (amostras grandes)

Considere uma população com média μ e σ^2. Suponha que temos uma amostra $X_1, \ldots, X_n$, então

$$\bar{X} \sim Normal\left(\mu, \frac{\sigma^2}{n}\right).$$

Propriedade IMPORTANTE da distribuição Normal

Se $x_1, \ldots, x_m$ valores observados da variável aleatória $X \sim N(\mu, \sigma^2)$, então $\bar{X} \sim Normal\left(\mu, \frac{\sigma^2}{n}\right)$.

Exemplo

Em um estudo da altura de pacientes, escolhemos 10 pacientes. Sabemos que a altura dos pacientes tem distribuição normal com média 185*cm* e desvio padrão 40*cm*. Qual a distribuição de $\bar{X}$? Qual a probabilidade da altura média dos pacientes escolhidos ser maior que a média populacional?

Solução:

- $\bar{X} \sim Normal\left(185, \frac{40}{10}\right)$, ou seja, $\bar{X} \sim Normal\,(185, 4)$.
- $$\begin{aligned} P(\bar{X} > 185) &= 1 - P(\bar{X} \leq 185) \\ &= 1 - \Phi\left(\frac{185 - 185}{2}\right) \\ &= 1 - \Phi(0) = 1 - 0,50 = 0,50 \end{aligned}$$

Exemplo

Considere uma variável aleatória discreta $X : \Omega \to \mathbb{R}$ que assume os valores $3, 6$, e 8 com, respectivamente, probabilidades $0,5$; $0,3$, e $0,2$. Uma amostra de 40 observações é sorteada, qual a probabilidade da média da amostra ser maior que 5?

Solução: Primeiramente, note que

$$\mu = 0,5 \cdot 3 + 0,3 \cdot 6 + 0,2 \cdot 8 = 4,9,$$
$$\sigma^2 = 0,5 \cdot (3 - 4,9)^2 + 0,3 \cdot (6 - 4,9)^2 + 0,2 \cdot (8 - 4,9)^2 = 4,09.$$

Usando o Teorema central do limite, temos que $\bar{X} \sim N\left(4,9; \frac{4,09}{40}\right)$ e

$$\begin{aligned} P(\bar{X} > 5) &= 1 - P(\bar{X} \leq 5) \\ &= 1 - \Phi\left(\frac{5 - 4,9}{\sqrt{\frac{4,09}{40}}}\right) \\ &= 1 - \Phi(0,32) = 1 - 0,6255 = 0,37. \end{aligned}$$

Exemplo

Distribuição Bernoulli

Lembre que $\mathrm{E}(X) = p$ e $\mathrm{Var}(X) = p(1-p)$.

Exemplo

Suponha que a prevalência do vírus HIV na África Subsariana é 10%. Um médico selecionou 40 pacientes desta região. Qual a probabilidade de no máximo 20% desses pacientes estarem infectados pelo vírus?

Solução: Pelo Teorema central do limite, temos que $\hat{p} \sim N\left(0,1; \dfrac{0,1 \cdot 0,9}{40}\right)$. Logo, temos que

$$P(\hat{p} < 0,2) = \Phi\left(\frac{0,2 - 0,1}{\sqrt{\frac{0,1 \cdot 0,9}{40}}}\right)$$
$$= \Phi(2,11) = 0,9826$$

Exemplo

Distribuição poison

Lembre que $E(X) = \lambda$ e $Var(X) = \lambda$.

Exemplo

A emissão de partículas radioativas alfa de um isótopo em um minuto é modelada através de um distribuição poison com média 5. Um físico analisar cinco amostras desse isótopo e observou o número de partículas alfa emitidas para uma amostra com observações x_1, x_2, x_3, x_4, x_5. Qual a probabilidade da média de partículas emitidas nessa amostra com cinco valores ser maior que seis?

Solução: Pelo teorema central do limite, temos que $\bar{X} \sim N\left(\lambda, \frac{\lambda}{5}\right)$. Então

$$
\begin{aligned}
P(\bar{X} > 6) &= 1 - P(\bar{X} \leq 6) \\
&= 1 - \Phi\left(\frac{6-5}{\sqrt{\frac{5}{5}}}\right) \\
&= 1 - \Phi(1) = 1 - 0,8413 = 0,1587
\end{aligned}
$$

Exemplo

Distribuição exponencial

Lembre que $\mathrm{E}(X) = \frac{1}{\alpha} = \mu$ e $\mathrm{Var}(X) = \frac{1}{\alpha^2} = \mu^2$.

Exemplo

Uma indústria fabrica lâmpadas especiais que ficam em operação continuamente. A fabricante afirma que as lâmpadas duram em média 8000 horas. Um órgão de controle teste 10 lâmpadas. Assumindo que o fabricante diz a verdade, qual a probabilidade do órgão regulador obter uma média de no máximo 7000 horas para a amostra?

Solução: Pelo teorema do limite central, temos que $\bar{X} \sim N\left(\mu; \dfrac{\mu^2}{n}\right)$. Então,

$$\begin{aligned} P(\bar{X} < 7000) &= \Phi\left(\frac{7000 - \mu}{\sqrt{\frac{\mu^2}{n}}}\right) = \Phi\left(\frac{7000 - 8000}{\sqrt{\frac{8000^2}{10}}}\right) \\ &= \Phi(-0,4) \\ &= 0,3446. \end{aligned}$$