Variáveis Aleatórias

Função de Distribuição de Probabilidade Acumulada

Definimos *função de distribuição de probabilidade acumulada* da variável aleatória X como sendo:

$$\boldsymbol{F(x) = P(X \leq x)}$$

Para o caso discreto, temos:

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} p(x_i)$$

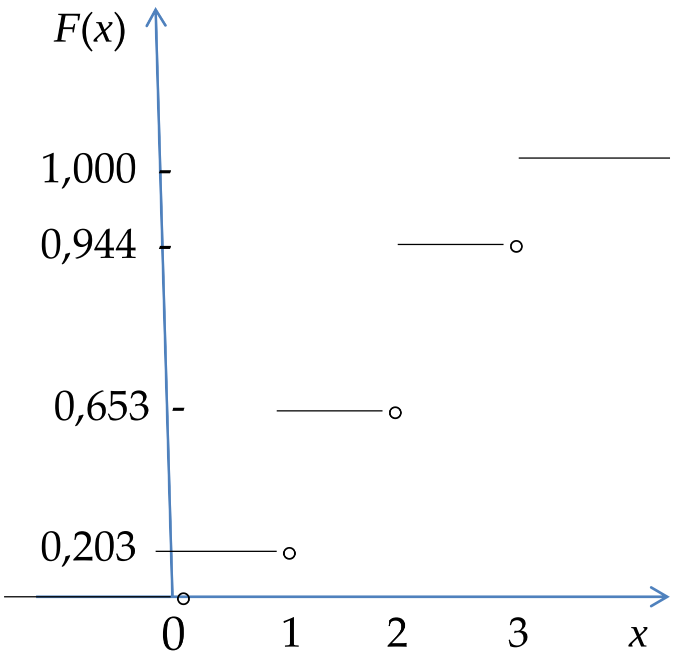
Exemplo anterior

$$F(0) = P(X \leq 0) = 0{,}203$$
$$F(1) = P(X \leq 1) = 0{,}653$$
$$F(2) = P(X \leq 2) = 0{,}944$$
$$F(3) = P(X \leq 3) = 1$$

Então,

$$F(x) = \begin{cases} 0, \text{ se } x < 0 \\ 0{,}203, \text{ se } 0 \leq x < 1 \\ 0{,}653\text{ , se } 1 \leq x < 2 \\ 0{,}944, \text{ se } 2 \leq x < 3 \\ 1, \text{ se } x \geq 3 \end{cases}$$

Representação gráfica de $F(x)$



Observações importantes de $F(x)$:

- O domínio de F é a reta real
- $\lim_{x \to -\infty} F(x) = 0$ e $\lim_{x \to +\infty} F(x) = 1$
- F é não decrescente, isto é, se $x_1 \le x_2$, tem-se $F(x_1) \le F(x_2)$
- A magnitude de cada salto é $p(x_i)$

Exercício

Considere uma variável aleatória com a seguinte função de probabilidade:

$$P(X=k)=\begin{cases} c, \text{ para } k=1, 3, 5 \\ 2c, \text{ para } k=2, 4 \end{cases}$$

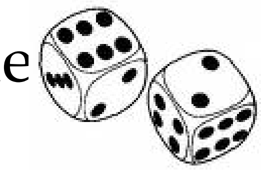
a) Determine o valor da constante "c" que torna legítima a função de probabilidade acima.

b) Determine a função de distribuição acumulada F e construa seu gráfico

c) Calcule a $P(X>1)$, $P(X\geq 3)$, $P(X\leq 4)$, $P(5/2<X\leq 5)$

Exemplo 2:

Um dado é lançado duas vezes, de forma independe



Qual é a probabilidade da *soma dos pontos nos dois lançamentos ser menor do que 6*?

$$\Omega = \{(1,1), (1,2), (1,3), (1,4), (1,5), (1,6),$$
$$(2,1), (2,2), (2,3), (2,4), (2,5), (2,6),$$
$$(3,1), (3,2), (3,3), (3,4), (3,5), (3,6),$$
$$(4,1), (4,2), (4,3), (4,4), (4,5), (4,6),$$
$$(5,1), (5,2), (5,3), (5,4), (5,5), (5,6),$$
$$(6,1), (6,2), (6,3), (6,4), (6,5), (6,6)\}.$$

Qual é a probabilidade de cada ponto w_i de Ω?

Admitindo-se que o dado seja perfeitamente homogêneo e sendo os lançamentos independentes,

$$P(w_i) = 1/36, \text{ qualquer } w_i \in \Omega.$$

Defina X : soma dos pontos nos dois lançamentos do dado.

$\Rightarrow$

Função de probabilidade de X:

x	2	3	4	5	6	7	8	9	10	11	12
$P(X=x)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Então,

$$P(X < 6) = P(X=5) + P(X=4) + P(X=3) + P(X=2)$$

$$= \frac{4}{36} + \frac{3}{36} + \frac{2}{36} + \frac{1}{36} = \frac{10}{36} = 0{,}278$$

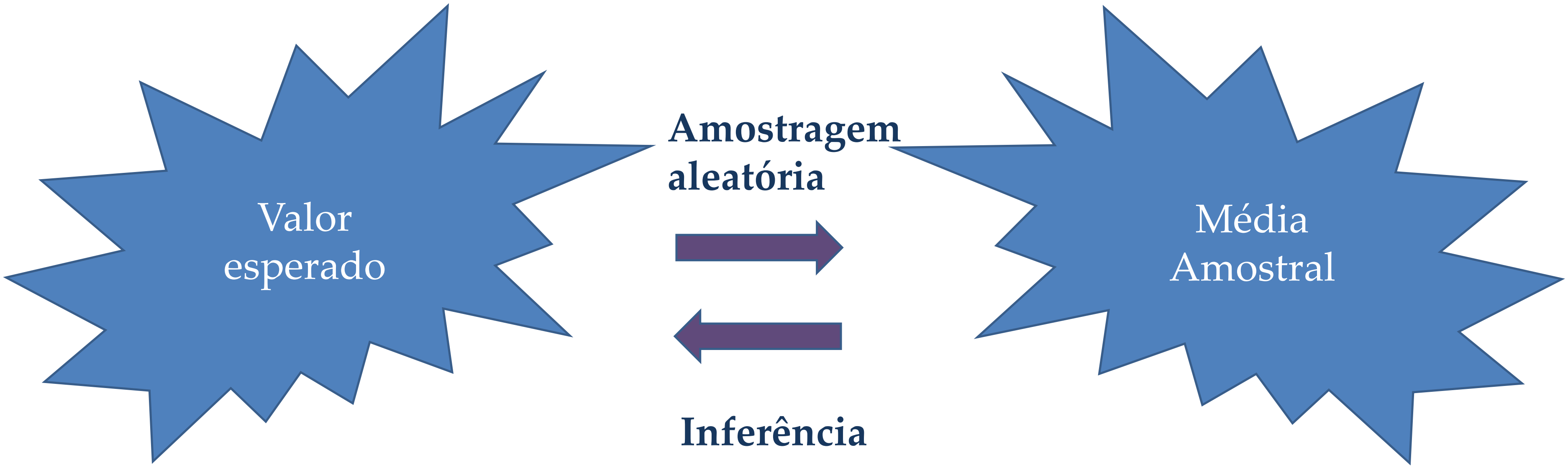
$\rightarrow$ Como seria a distribuição de probabilidades de Y : valor máximo obtido dentre os dois lançamentos do dado?

Na tentativa de resumir o comportamento de uma variável aleatória vamos estudar uma medida que estuda a tendência central da variável aleatória, chamada de *esperança* ou *valor esperado de uma variável aleatória*. Para ilustrar os vários momentos em que lidamos com o valor esperado vamos apresentar alguns exemplos.

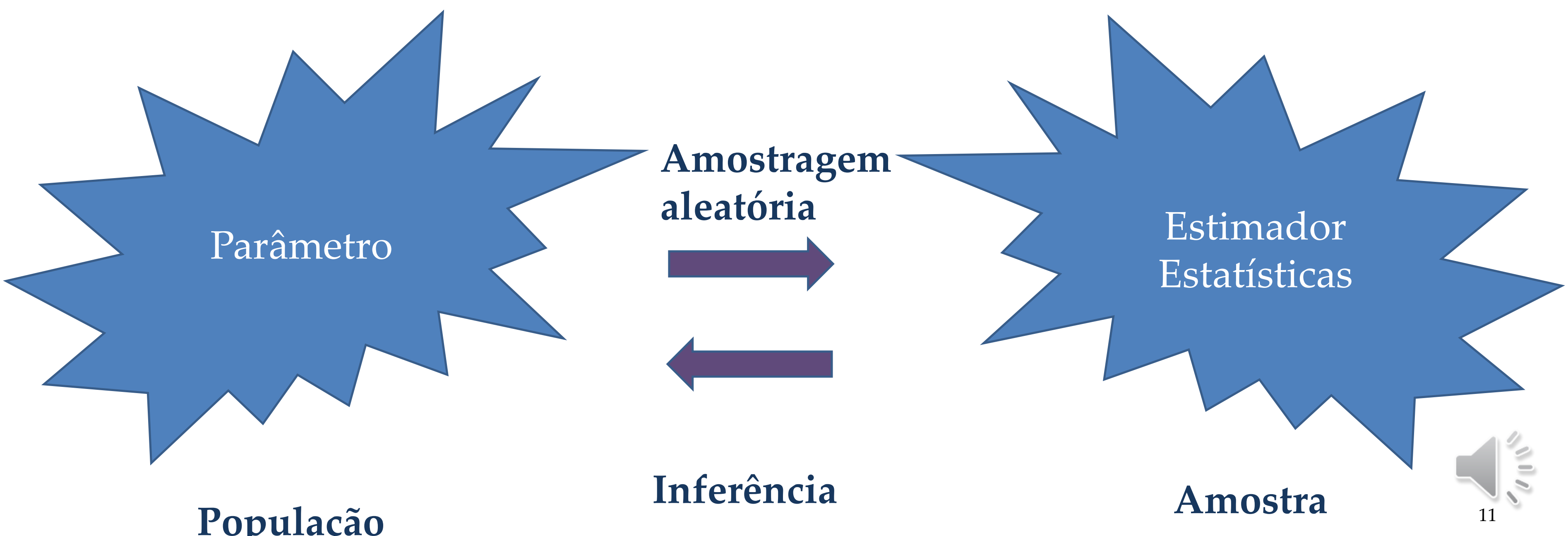
Amostragem

Número de filhos que a mulher no Brasil tem

Número de filhos que algumas mulheres no Brasil tem



A inferência estatística a respeito de um **parâmetro** é baseada na informação disponível em alguma amostra retirada da população de interesse. Geralmente, isto é feito utilizando **estatísticas**, que são **funções da amostra**. Por exemplo, a média amostral é uma estatística. As estatísticas utilizadas para estimar um parâmetro populacional têm uma distribuição de probabilidade.



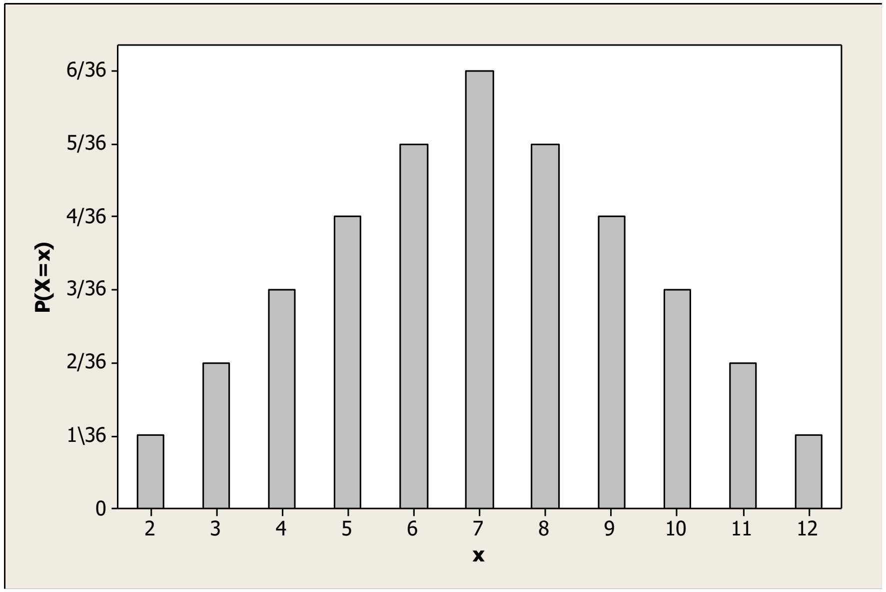
Valor Esperado e Variância

No Exemplo 2: *Qual é o valor médio da soma dos pontos (X) no lançamento de dois dados?*

Ω = {(1,1), (1,2), (1,3), (1,4), (1,5), (1,6),
(2,1), (2,2), (2,3), (2,4), (2,5), (2,6),
(3,1), (3,2), (3,3), (3,4), (3,5), (3,6),
(4,1), (4,2), (4,3), (4,4), (4,5), (4,6),
(5,1), (5,2), (5,3), (5,4), (5,5), (5,6),
(6,1), (6,2), (6,3), (6,4), (6,5), (6,6)}

$\Rightarrow$ 36 pontos igualmente prováveis

x	$P(X = x)$
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	3/36
12	1/36



Definição: **Valor Esperado ("média")**

Dada a *v.a.* X, assumindo os valores x_1, x_2, ..., x_n, chamamos de *valor médio*, ou *valor esperado*, ou *esperança matemática* da distribuição de X, o valor

$$E(X) = x_1 P(X = x_1) + \cdots + x_n P(X = x_n) = \sum_{i=1}^{n} x_i \, P(X = x_i)$$

Notação: $\boldsymbol{\mu = E(X)}$

No Exemplo 2, para média de X (soma de pontos), temos:

$$E(X) = 2 \times \frac{1}{36} + 3 \times \frac{2}{36} + \cdots + 11 \times \frac{2}{36} + 12 \times \frac{1}{36} = \frac{252}{36} = 7,$$

ou seja, em média, a soma dos pontos no lançamento dos dois dados é igual a 7.

Definição: **Variância**

É o valor esperado da *v.a.* $(X - E(X))^2$, ou seja, se X assume os valores $x_1, x_2, \ldots, x_n$, então

$$Var(X) = E\{[X - E(X)]^2\}$$

$$= \sum_{i=1}^{n} [x_i - E(X)]^2 \times P(X = x_i)$$

Desenvolvendo a fórmula acima, e lembrando que $E(X) = \mu$, obtemos a seguinte fórmula alternativa:

$$Var(X) = E(X^2) - [E(X)]^2$$
$$= E(X^2) - \mu^2$$

$$= \sum_{i=1}^{n} x_i^2 \times P(X = x_i) - \mu^2$$

A notação usual de variância é

$$\boldsymbol{Var(X) = \sigma^2}$$

O Desvio Padrão é definido como a raiz quadrada positiva da variância, isto é,

$$DP(X) = \sqrt{Var(X)}$$

Notação: $\boldsymbol{DP\ (X) = \sigma}$

No Exemplo 2,

x	2	3	4	5	6	7	8	9	10	11	12
$P(X=x)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

$$Var(X) = (2-7)^2 \times \frac{1}{36} + (3-7)^2 \times \frac{2}{36} + \ldots + (11-7)^2 \times \frac{2}{36} + (12-7)^2 \times \frac{2}{36}$$

$$= \frac{210}{36} = 5{,}83$$

Podemos também calcular pela fórmula alternativa

$$E(X^2) = 2^2 \times \frac{1}{36} + 3^2 \times \frac{2}{36} + \ldots + 11^2 \times \frac{2}{36} + 12^2 \times \frac{2}{36}$$

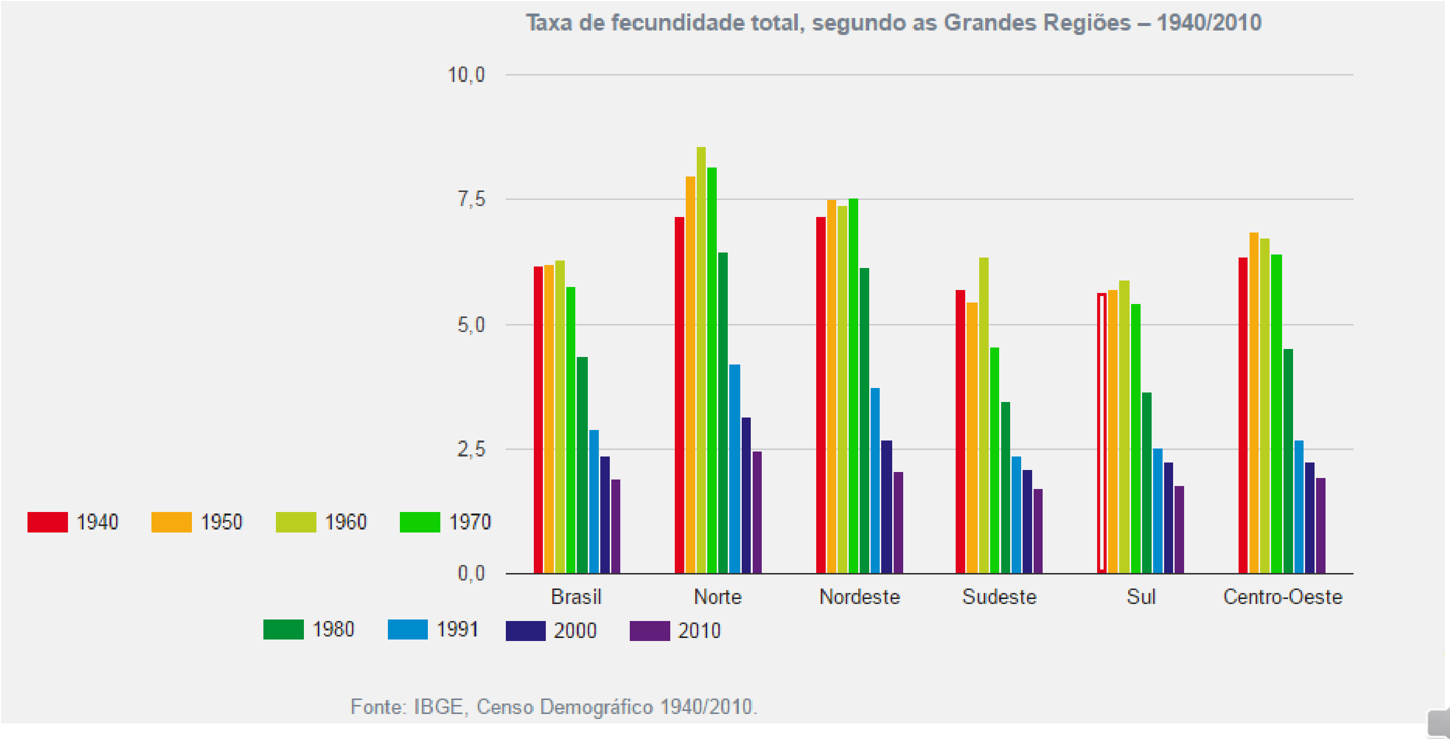
$$= \frac{1974}{36} = 54{,}83$$

e, portanto, $Var(X) = E(X^2) - (E(X))^2 = 54{,}83 - 7^2 = 5{,}83.$

Calcule o número esperado de filhos de uma mulher do Brasil

A taxa de fecundidade indica quantos filhos, em média, tem a mulher brasileira. No Brasil, segundo o Censo 2010, as mulheres têm, em média, 1,9 filho. Como este número é uma média, existem mulheres com um filho e mulheres com dois, três ou mais filhos.

O número de filhos por mulher vem se reduzindo no Brasil desde a década de 1960. Uma redução que vem ocorrendo em todas as regiões brasileiras. Veja o gráfico a seguir:



http://7a12.ibge.gov.br/vamos-conhecer-o-brasil/nosso-povo/nupcialidade-e-fecundidade.html

Se saber o número de filhos de uma mulher "qualquer" de seu bairro?
Que representa esse valor?
Se considerar várias amostras que acontece com estes valores?

Propriedades:

1) Se $P(X = a) = 1$ (i.é., a variável é uma constante), então

$$E(X) = a \quad \text{e} \quad Var(X) = 0.$$

2) Se $Y = a + bX$, em que a e b são constantes, então

$$E(Y) = E(a + bX) = a + bE(X)$$

e

$$Var(Y) = Var(a + bX) = b^2 \, Var(X).$$

Exemplo

Suponha que o professor da disciplina de Estatística, ao fim do curso, para os alunos que não compareceram em suas aulas elaborou uma prova de múltipla escolha, consistindo de 10 questões, cada uma com 5 alternativas, sendo somente uma correta.

Suponha que nenhum desses alunos estudou para essa prova, acreditando que, por ser prova de múltipla escolha, se sairiam bem.

O professor estabeleceu que para aprovar na disciplina deve responder corretamente ao menos 6 questões. Sorteando aleatoriamente um aluno dentre os que farão essa prova, qual é a probabilidade de ele ser aprovado?

Em média, quantas questões um aluno acertará?

Solução:

Considere a *v.a.* X: número de questões respondidas corretamente pelo aluno selecionando, dentre as 10 questões.

Então o evento de interesse,

Sucesso (S) é "questão respondida corretamente", com $P(S) = p$ = ?

e Fracasso (F), "questão respondida incorretamente", com $P(F)$= ?

Resultado:

Se $X \sim b(n; p)$, então

$$\text{valor esperado: } \mu = E(X) = n \times p$$

$$\text{variância: } \sigma^2 = Var(X) = n \times p \times (1 - p)$$

No Exemplo,

espera-se que o aluno sorteado acerte, em média, $10 \times 0{,}2 = 2$ questões.

Ainda no Exemplo,

supondo que 40 alunos farão essa prova de múltipla escolha, quantos alunos o professor espera que sejam aprovados?

Sendo X a *v.a.* que representa o número de alunos que serão aprovados na prova, dentre os 40 que irão fazê-la,

então $X \sim b(40; 0{,}00637)$.

Logo, em média, $40 \times 0{,}00637 = 0{,}25$ alunos deverão ser aprovados, dentre os 40 que farão essa prova.

Exemplo

O escore em um teste internacional de proficiência na língua inglesa varia de 0 a 700 pontos, com mais pontos indicando um melhor desempenho. Informações, coletadas durante vários anos, permitem estabelecer o seguinte modelo de probabilidade para o desempenho no teste:

Pontos	**[0,200)**	**[200,300)**	**[300,400)**	**[400,500)**	**[500,600)**	**[600,700]**
p_i	0,06	0,15	0,16	0,25	0,27	0,11

Várias universidades americanas, exigem um escore mínimo de 600 pontos para aceitar candidatos de países de língua não inglesa. De um grande grupo de estudantes que prestaram o último exame, escolhemos, ao acaso, 10 deles.

Qual é a probabilidade de no máximo 2 atenderem ao requisito mencionado? Em um grupo de 2200 candidatos espera-se que quantos sejam aprovados?