UNIDADE III - Elementos de probabilidades

3.1. Introdução à teoria das probabilidades
   3.1.1. Introdução
   3.1.2. Conceitos fundamentais
   3.1.3. Conceitos de probabilidade
   3.1.4. Teoremas para o cálculo de probabilidades
   3.1.5. Probabilidade condicional e independência
3.2. Variáveis aleatórias
   3.2.1. Introdução e conceito
   3.2.2. Variáveis aleatórias discretas
   3.2.3. Variáveis aleatórias contínuas
3.3. Distribuições de probabilidade
   3.3.1. Distribuições de probabilidade de variáveis discretas
   3.3.2. Distribuições de probabilidade de variáveis contínuas

Profª. Clause Piana

3.3 Distribuições de probabilidade

- Introdução
- Resumo das distribuições discretas
  - Principais características
  - Exemplo
- Aproximações entre as distribuições

Introdução

Por que a variável é denominada aleatória?

Porque seus valores ocorrem de forma aleatória ou probabilística.

Uma das mais importantes maneiras pelas quais a aleatoriedade nos afeta é por meio de sua influência nas medições.

Por exemplo, quando um professor avalia um trabalho de um aluno a nota não é uma descrição do grau de qualidade do trabalho, mas uma **medição** dessa qualidade. Nesse caso, o instrumento de medição é o professor, e a avaliação de tal profissional, como qualquer medição, está sujeita a variações e erros aleatórios.

A incerteza da medição é ainda mais problemática quando a quantidade medida é subjetiva.

Para entender as medições é fundamental conhecer a **forma como os dados estão distribuídos**.

As **medidas** descritivas são uma ferramenta essencial para obter essa informação.

Outro exemplo: suponha que 15 enólogos vão avaliar um vinho.

Situação 1. Os 15 enólogos concordam que a nota do vinho é 90.

Situação 2. Os enólogos expressam as notas 80, 81, 82, 87, 89, 89, 90, 90, 90, 91, 91, 94, 97, 99 e 100.

Uma das maneiras de gerar um número único a partir de uma série de medições discordantes é calcular a **média**. Assim, podemos resumir as opiniões utilizando a média das notas atribuídas ao vinho. Mas essa informação é suficiente?

Os dois conjuntos de dados tem a mesma média, mas diferem no quanto variam a partir dessa média.

O **desvio padrão** caracteriza o quanto um conjunto de dados se aproxima da média, o que também pode significar a **incerteza dos dados.**

Na situação 2, o desvio padrão é 6 e o que podemos realmente dizer sobre o vinho é que sua nota provavelmente se situe entre 84 e 96.

Notamos assim que a aleatoriedade (os erros aleatórios) gera variabilidade.

A Estatística busca a regularidade presente na variabilidade.

Esta regularidade permite que as variáveis aleatórias sejam representadas por modelos matemáticos, que são denominados distribuições de probabilidade.

As distribuições de probabilidade constituem a espinha dorsal da metodologia estatística, pois é com base nesses modelos que a Estatística cria técnicas que possibilitam tomar decisões válidas apesar da variabilidade e na presença de incerteza.

Na pesquisa científica o pesquisador precisa tomar decisões e obter conclusões na presença de variabilidade.

Distribuições de probabilidade

O que é uma distribuição de probabilidade?

Uma distribuição de probabilidade é essencialmente um **modelo de descrição probabilística de uma população**.

⇨ **População estatística** é o conjunto de todos os valores de uma variável aleatória.

Os modelos têm aplicações gerais e são individualizados através dos parâmetros.

⇨ **Parâmetros:** caracterizações numéricas que permitem a individualização de um modelo em determinado contexto

⇨ No estudo de uma variável aleatória é importante saber:

1. O tipo de distribuição que é determinado pela:

- **função de probabilidade**, se a variável é **discreta**

$$P(X=x)=\frac{C_7^x\,C_3^{3-x}}{C_{10}^3}\text{, para } S_X = \{0, 1, 2, 3\}$$

- **função densidade de probabilidade**, se a variável é **contínua**

$$f(x)=2e^{-2x},\ \text{para } S_X=[0, \infty)$$

2. Os parâmetros da distribuição

3. As medidas descritivas da distribuição (média, variância, assimetria)

Distribuições discretas

1. **Distribuição de Bernoulli**
2. **Distribuição Binomial**
3. **Distribuição Hipergeométrica**
4. **Distribuição de Poisson**

X é o número sucessos

5. Distribuição Multinomial
6. Distribuição Geométrica
7. Distribuição Binomial Negativa
8. Distribuição Hipergeométrica Negativa
9. Distribuição Uniforme

Resumo das distribuições discretas

1. Distribuição de Bernoulli

Descreve probabilisticamente experimentos aleatórios que possuem apenas dois resultados possíveis.

X= número de sucessos $\quad S_X = \{0, 1\}$

Função de probabilidade

$$P(X = x) = \pi^x (1 - \pi)^{1-x}$$

Parâmetro: π = probabilidade de sucesso

Medidas descritivas

$E(X) = \mu = \pi$

$V(X) = \sigma^2 = \pi\,(1- \pi)$

$$a_3 = \frac{(1-\pi) - \pi}{\sqrt{\pi(1-\pi)}}$$

Exemplo:

Um atirador dispara um tiro contra um alvo.

S = {acertar, errar}

Consideramos um dos resultados como sucesso.

sucesso = acertar

fracasso = errar

X = número de sucessos

$S_X = \{0, 1\}$

π = probabilidade de sucesso

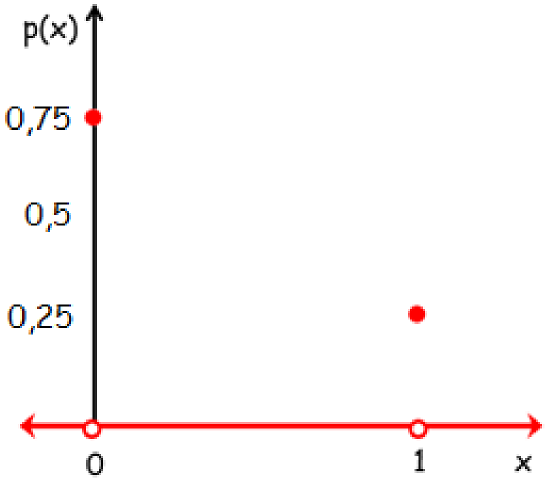
$1-\pi$ = probabilidade de fracasso

Exemplo:

Um atirador dispara um tiro contra um alvo. Sabe-se que a probabilidade do atirador acertar o alvo é 0,25.

X = número de sucessos (acertos) $S_X = \{0, 1\}$

$\pi = 0,25$



Medidas descritivas

$E(X) = \mu = \pi = 0,25$

$V(X) = \sigma^2 = \pi\,(1-\pi) = 0,25 . 0,75 = 0,1875$

$$a_3 = \frac{(1-\pi)-\pi}{\sqrt{\pi(1-\pi)}} = \frac{0,75-0,25}{\sqrt{0,25 . 0,75}} = 1,155$$

X ~ Ber (π=0,25)

2. Distribuição binomial

Descrição probabilística de uma sequência de experimentos de Bernoulli **independentes**.
Importante no contexto de amostragem **com reposição**.

X= número de sucessos $S_X = \{0, 1, ..., n\}$

Função de probabilidade

$$P(X = x) = P_n^{x,n-x} \, \pi^x \, (1-\pi)^{n-x}$$

Parâmetros: n = número de repetições do experimento
π = probabilidade de sucesso

Medidas descritivas

$E(X) = \mu = n\pi$

$V(X) = \sigma^2 = n\pi(1-\pi)$

$$a_3 = \frac{(1-\pi)-\pi}{\sqrt{n\pi(1-\pi)}}$$

Exemplo:

Quatro atiradores disparam contra um alvo. Todos foram igualmente treinados e têm a mesma probabilidade de acertar o alvo, que é 0,25.

S = {AAAA, AAAE, AAEA, AEAA, EAAA, AAEE, ..., EEEE} $\#S = 2^4 = 16$

X = número de sucessos

$S_X = \{0, 1, 2, 3, 4\}$

n = número de repetições do experimento de Bernoulli

π = probabilidade de sucesso

n = 4

π = 0,25

$P(X=4) = 0{,}25^4$

$P(X=3) = 4 \times 0{,}25^3 \times 0{,}75^1$

$P(X=2) = 6 \times 0{,}25^2 \times 0{,}75^2$

$$P(X = x) = \underbrace{P_n^{x,n-x}}_{\text{Número de casos}} \underbrace{\pi^x (1-\pi)^{n-x}}_{\text{Probabilidade de um caso}}$$

Exemplo:

Quatro atiradores disparam contra um alvo. Todos foram igualmente treinados e têm a mesma probabilidade de acertar o alvo, que é 0,25.

S = {AAAA, AAAE, AAEA, AEAA, EAAA, AAEE, ..., EEEE} $\#S = 2^4 = 16$

X = número de sucessos

S_X = {0, 1, 2, 3, 4}

n = número de repetições do experimento de Bernoulli

π = probabilidade de sucesso

n = 4

π = 0,25

$P(X=4) = 0{,}25^4$

$P(X=3) = 4 \times 0{,}25^3 \times 0{,}75^1$

$P(X=2) = 6 \times 0{,}25^2 \times 0{,}75^2$

$$P(X = x) = P_n^{x,n-x}\,\pi^x(1-\pi)^{n-x}$$

$$P_n^{x,n-x} = \frac{n!}{x!(n-x)!}$$

Exemplo:

Quatro atiradores disparam contra um alvo. Todos foram igualmente treinados e têm a mesma probabilidade de acertar o alvo, que é 0,25.

X = número de sucessos (acertos) $S_X = \{0, 1, 2, 3, 4\}$

$n = 4$

$\pi = 0,25$

Função de probabilidade

$$P(X = x) = P_4^{x,4-x}\, 0,25^x\, (1-0,25)^{4-x}, \text{ para } S_X = \{0, 1, 2, 3, 4\}$$

$$P(X = 0) = P_4^{0,4}\, 0,25^0\, (1-0,25)^4$$

$$P(X = 1) = P_4^{1,3}\, 0,25^1\, (1-0,25)^3$$

...

$$P(X = 4) = P_4^{4,0}\, 0,25^4\, (1-0,25)^0$$

Exemplo:

Quatro atiradores disparam contra um alvo. Todos foram igualmente treinados e têm a mesma probabilidade de acertar o alvo, que é 0,25.

X = número de sucessos (acertos) $S_X = \{0, 1, 2, 3, 4\}$

$n = 4$ $\pi = 0,25$

Medidas descritivas

$$E(X) = \mu = n\pi = 4.0,25 = 1$$

$$X \sim Bin\ (n=4,\ \pi=0,25)$$

$$V(X) = \sigma^2 = n\pi(1-\pi) = 4.0,25.0,75 = 0,75$$

$$a_3 = \frac{(1-\pi)-\pi}{\sqrt{n\pi(1-\pi)}} = \frac{0,75-0,25}{\sqrt{4.0,25.0,75}} = 0,433$$

Distribuição binomial

$$P(X = x) = P_n^{x,n-x}\, \pi^x\, (1 - \pi)^{n-x} \quad \text{, para } S_X = \{0, 1, \ldots, n\}$$

$n = 1$

$$P(X = x) = P_1^{x,1-x}\, \pi^x (1 - \pi)^{1-x}$$

$$P(X = x) = \pi^x (1 - \pi)^{1-x}$$

Distribuição de Bernoulli

$$P(X = x) = \pi^x\, (1 - \pi)^{1-x}$$

3. Distribuição hipergeométrica

Descrição probabilística de uma sequência de experimentos de Bernoulli **dependentes**.
Importante no contexto de amostragem **sem reposição**.

X= número de sucessos $S_X=\{\max(0, n-N_2), ..., \min(n,N_1)\}$

Função de probabilidade $S_X=\{0, 1, 2, ..., n\}$

$$P(X=x)=\frac{C_{N_1}^{x}\,C_{N_2}^{n-x}}{C_{N}^{n}}$$

Parâmetros: n = número de repetições do experimento
N = tamanho da população
N_1 = tamanho da subpopulação de interesse

Medidas descritivas

$$E(X)=\mu=n\frac{N_1}{N} \qquad V(X)=\sigma^2=n\frac{N_1}{N}\frac{N_2}{N}\left(\frac{N-n}{N-1}\right)$$

Exemplo:

Uma bandeja contém 10 xícaras de cafezinho, sete com açúcar e três sem açúcar. Três pessoas se aproximam da bandeja e se servem.

$S = \{c_1c_2c_3, c_1c_2c_4, \ldots, c_1c_2s_1, \ldots, s_1s_2s_3\}$ $\#S = C_{10}^{3} = 120$

Consideramos um dos resultados como sucesso

sucesso = com açúcar
fracasso = sem açúcar

X = número de sucessos (com açúcar)

$S_X = \{0, 1, 2, 3\}$

n = número de repetições do experimento
N = tamanho da população ($N = N_1 + N_2$)
N_1 = tamanho da subpopulação de interesse (sucesso)

Exemplo:

Uma bandeja contém 10 xícaras de cafezinho, sete com açúcar e três sem açúcar. Três pessoas se aproximam da bandeja e se servem.

$S = \{c_1c_2c_3,\ c_1c_2c_4, \ldots,\ c_1c_2s_1,\ \ldots,\ s_1s_2s_3\}$ $\quad \#S = C_{10}^{3} = 120$

Consideramos um dos resultados como sucesso

sucesso = com açúcar
fracasso = sem açúcar

$n = 3$
$N = 10$
$N_1 = 7$
$N_2 = 3$

X = número de sucessos (com açúcar)

$S_X = \{0, 1, 2, 3\}$

n = número de repetições do experimento
N = tamanho da população ($N = N_1 + N_2$)
N_1 = tamanho da subpopulação de interesse (sucesso)

Exemplo:

Uma bandeja contém 10 xícaras de cafezinho, sete com açúcar e três sem açúcar. Três pessoas se aproximam da bandeja e se servem.

X = número de sucessos (açúcar) $S_X = \{0, 1, 2, 3\}$

$n = 3,\ N = 10,\ N_1 = 7$

Função de probabilidade

$$P(X=x) = \frac{C_7^x\, C_3^{3-x}}{C_{10}^3}$$, para $S_X = \{0, 1, 2, 3\}$

$$P(X=0) = \frac{C_7^0\, C_3^3}{C_{10}^3} \qquad P(X=2) = \frac{C_7^2\, C_3^1}{C_{10}^3}$$

$$P(X=1) = \frac{C_7^1\, C_3^2}{C_{10}^3} \qquad P(X=3) = \frac{C_7^3\, C_3^0}{C_{10}^3}$$

Exemplo:

Uma bandeja contém 10 xícaras de cafezinho, sete com açúcar e três com adoçante. Três pessoas se aproximam da bandeja e se servem.

X = número de sucessos (açúcar) $S_X = \{0, 1, 2, 3\}$

$n = 3, \; N = 10, \; N_1 = 7$

$N = N_1 + N_2$

$X \sim Hip\,(n=3, N=10, N_1=7)$

Medidas descritivas

$$E(X) = n\,\frac{N_1}{N} = 3\,\frac{7}{10} = 2{,}1$$

$$V(X) = n\,\frac{N_1}{N}\,\frac{N_2}{N}\left(\frac{N-n}{N-1}\right) = 3\,\frac{7}{10}\,\frac{3}{10}\left(\frac{10-3}{10-1}\right) = 0{,}49$$

Exemplo:

Uma bandeja contém 10 xícaras de cafezinho, sete com açúcar e três com adoçante. Três pessoas se aproximam da bandeja e se servem.

X = número de sucessos (açúcar) $S_X = \{0, 1, 2, 3\}$

$n = 3,\ N = 10,\ N_1 = 7$

$N = N_1 + N_2$

$X \sim Hip\,(n=3, N=10, N_1=7)$

Medidas descritivas

$$E(X) = n\frac{N_1}{N} = 3\frac{7}{10} = 2{,}1$$

$$V(X) = n\frac{N_1}{N}\frac{N_2}{N}\left(\frac{N-n}{N-1}\right) = 3\frac{7}{10}\frac{3}{10}\left(\frac{10-3}{10-1}\right) = 0{,}49$$

4. Distribuição de Poisson

Descrição probabilística da sequência de um **grande número** de fenômenos **independentes**, com probabilidade de sucesso **muito pequena**, que ocorrem por unidade de **tempo** ou de **superfície** ou de **volume**, com média de sucessos constante.
Importante na descrição de **eventos de ocorrência rara**.

X= número de sucessos $S_X=\{0, 1, 2, ...\}$ $\leftarrow$ espaço amostral infinito

Função de probabilidade

$$P(X=x)=e^{-\lambda}\frac{\lambda^x}{x!}$$

Parâmetro: λ = número médio de sucessos

Medidas descritivas

$$E(X) = \mu = \lambda \qquad a_3=\frac{1}{\sqrt{\lambda}}$$

$$V(X) = \sigma^2 = \lambda \qquad a_4=\sqrt{\lambda}$$

Exemplo:

A taxa média de chegada de clientes em um posto de serviços é de 0,5 por minuto.

X = número de clientes que chegam no posto de serviço por minuto

$S_X = \{0, 1, 2, 3, ...\}$

λ = número médio de sucessos (clientes)

$\lambda = E(X) = 0,5$

Função de probabilidade

$$P(X = x) = e^{-0,5}\frac{0,5^x}{x!}$$, para $S_X = \{0, 1, 2, 3, ...\}$

Exemplo:

A taxa média de chegada de clientes em um posto de serviços é de 0,5 por minuto.

X = número de clientes que chegam no posto de serviço por minuto

$S_X = \{0, 1, 2, 3, ...\}$ $\lambda = 0{,}5$

Função de probabilidade

$$P(X = x) = e^{-0{,}5}\frac{0{,}5^x}{x!}$$, para $S_X = \{0, 1, 2, 3, ...\}$

$$P(X = 0) = e^{-0{,}5}\frac{0{,}5^x}{0!} = 0{,}6065$$

$$P(X = 2) = e^{-0{,}5}\frac{0{,}5^x}{2!} = 0{,}07581$$

$$P(X = 1) = e^{-0{,}5}\frac{0{,}5^x}{1!} = 0{,}3033$$

$$P(X = 3) = e^{-0{,}5}\frac{0{,}5^x}{3!} = 0{,}01263$$

Exemplo:

A taxa média de chegada de clientes em um posto de serviços é de 0,5 por minuto.

X = número de clientes que chegam no posto de serviço por minuto

$S_X = \{0, 1, 2, 3, ...\}$ $\lambda = 0{,}5$

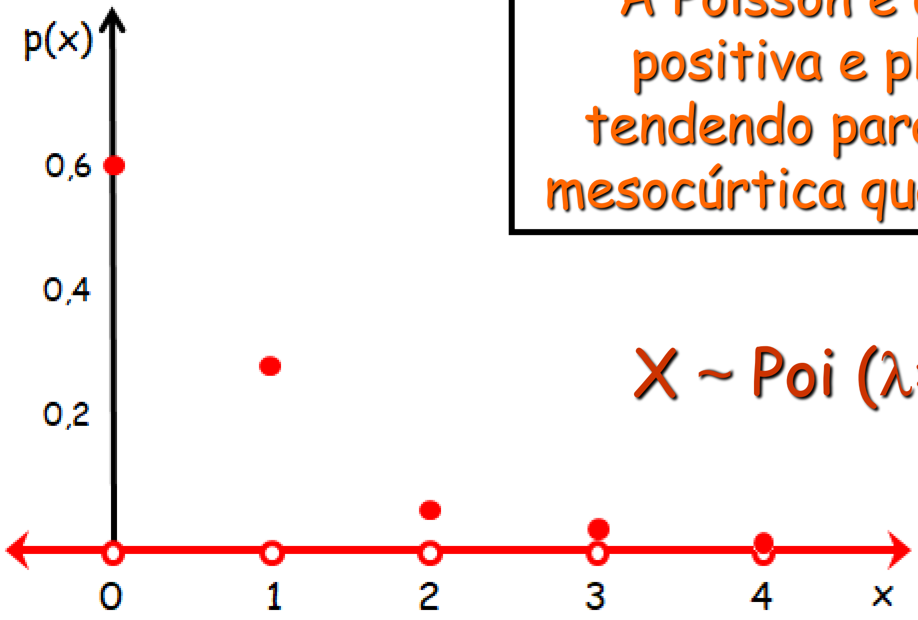
Medidas descritivas

$E(X) = \mu = \lambda = 0{,}5$

$V(X) = \sigma^2 = \lambda = 0{,}5$

$a_3 = \frac{1}{\sqrt{\lambda}} = 1{,}414$

$a_4 = \sqrt{\lambda} = 0{,}7071$



A Poisson é assimétrica positiva e platicúrtica, tendendo para simétrica e mesocúrtica quando μ cresce.

$X \sim Poi\ (\lambda=0{,}5)$

Distribuições discretas

1. Distribuição de Bernoulli

$$P(X=x)=\pi^{x}(1-\pi)^{1-x}$$

X ~ Ber (π=0,25)

2. Distribuição Binomial

$$P(X=x)=P_{n}^{x,n-x}\,\pi^{x}(1-\pi)^{n-x}$$

X ~ Bin (n=4, π=0,25)

3. Distribuição Hipergeométrica

$$P(X=x)=\frac{C_{N_1}^{x}\,C_{N_2}^{n-x}}{C_{N}^{n}}$$

X ~ Hip (n=3, N=10, N_1=7)

4. Distribuição de Poisson

$$P(X=x)=e^{-\lambda}\frac{\lambda^{x}}{x!}$$

X ~ Poi (λ=0,5)

Aproximações entre as distribuições

Formas limites da distribuição binomial

- Em determinadas circunstâncias, uma distribuição de probabilidade pode tender para outra.

- Os casos mais importantes de aproximações entre distribuições são:

  1. **Hipergeométrica $\rightarrow$ Binomial**
  2. **Binomial $\rightarrow$ Poisson**
  3. **Binomial $\rightarrow$ Normal**

Formas limites da distribuição binomial

1. Hipergeométrica → Binomial

Quando N (tamanho da população) é muito grande (ou tende para $+\infty$), a distribuição **hipergeométrica** se aproxima da distribuição **binomial**.

fator de correção

Binomial	Hipergeométrica
$V(X) = n\pi(1 - \pi)$	$V(X) = n \frac{N_1}{N} \frac{N_2}{N} \left(\frac{N-n}{N-1}\right)$

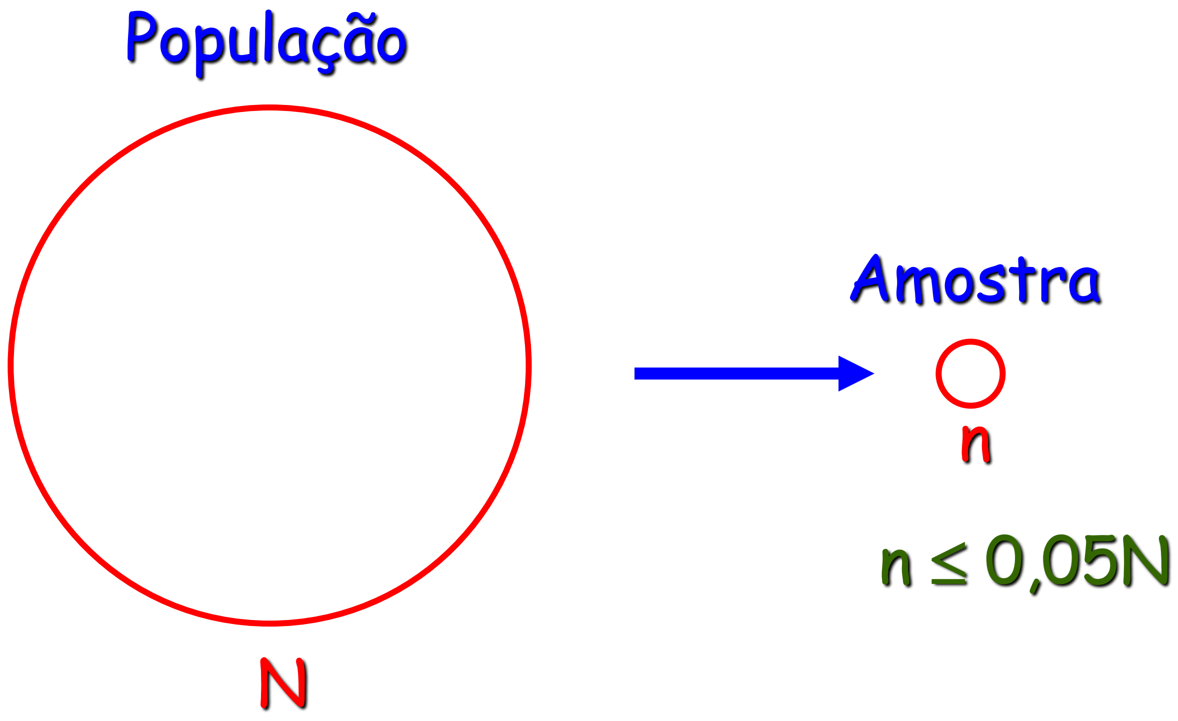
Quando N tende para infinito, o fator de correção se aproxima de 1.

tende a zero

$$\frac{N-n}{N-1} \cong \frac{N-n}{N} = \frac{N}{N} - \frac{n}{N} = 1 - \frac{n}{N} \cong 1$$

tende a $+\infty$

Esta aproximação é considerada satisfatória quando o número de elementos retirados (n) não excede 5% da população (N), isto é, $n \leq 0{,}05N$.



Quando o tamanho da amostra representa menos de 5% do tamanho da população, a população é tão grande em relação a amostra que pode ser considerada infinita.

2. Binomial $\rightarrow$ Poisson

Quando n (número de repetições do experimento) é muito grande (ou tende para $+\infty$) e π (probabilidade de sucesso) é muito pequena (ou tende para 0) a distribuição **binomial** se aproxima da distribuição de **Poisson**.

$$n\pi < 10 \quad \text{e} \quad n \geq 100$$

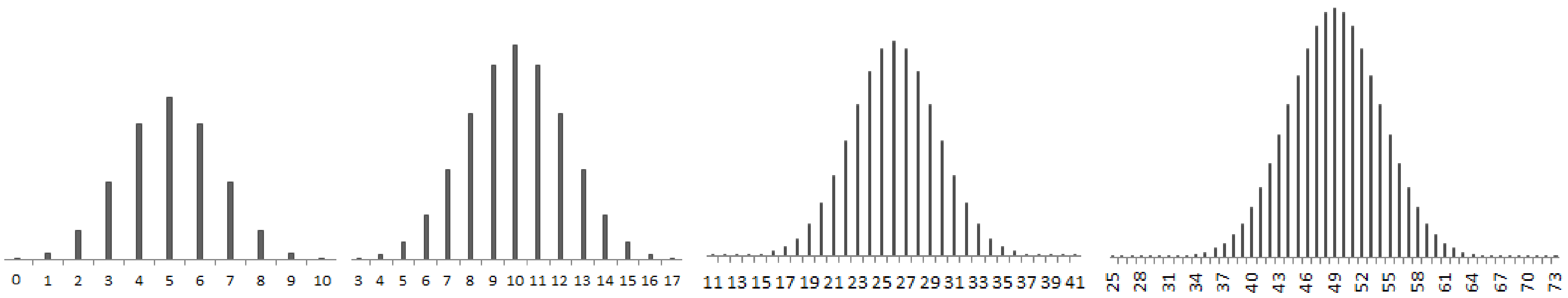
E(X)

3. Binomial → Normal

Quando n (número de repetições do experimento) é grande (ou tende para +∞) e π (probabilidade de sucesso) se aproxima de 0,5, a distribuição **binomial** se aproxima da distribuição **normal**.

Se **$\pi=(1-\pi)=0,5$**, a distribuição binomial será simétrica.

A distribuição binomial se aproxima da normal à medida que o n cresce.



n = 10 n = 20 n = 50 n = 100

Bibliografia

BUSSAB, W.O.; MORETTIN, P.A. **Estatística Básica**. São Paulo: Saraiva. 2006. 526p.

FERREIRA, D.F. **Estatística Básica**. Lavras: Editora UFLA, 2005, 664p.

FREUND, J.E., SIMON, G.A. **Estatística Aplicada. Economia, Administração e Contabilidade**. 9.ed., Porto Alegre: Bookman, 2000. 404p.

MEYER, P. L. **Probabilidade: aplicações à estatística**. Rio de Janeiro: LTC, 1976.

MLODINOW, L. **O andar do bêbado. Como o acaso determina nossas vidas**. Rio de Janeiro: Editora Zahar, 2009, 264p.

SILVEIRA JÚNIOR, P., MACHADO, A.A., ZONTA, E.P., SILVA, J.B. da **Curso de Estatística**. v.2, Pelotas: Universidade Federal de Pelotas, 1992. 234p.

Sistema Galileu de Educação Estatística. Disponível em: http://www.galileu.esalq.usp.br/topico.html