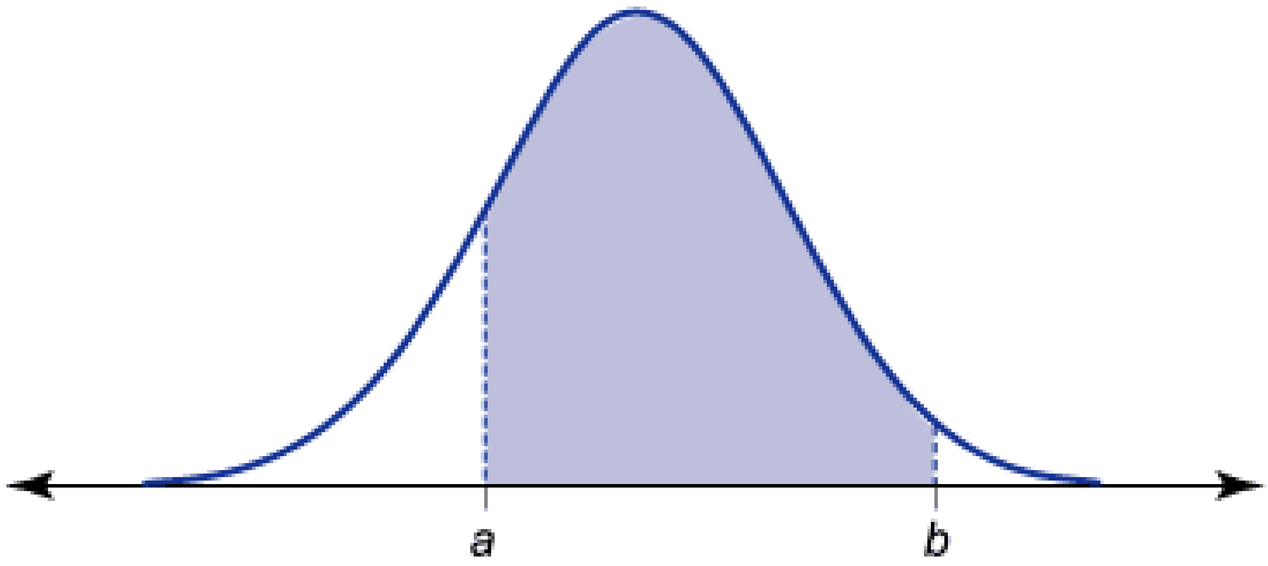


Estimação de Parâmetros

Esperança (Valor esperado)

- Representa o valor médio "esperado" de uma experiência se ela for repetida muitas vezes

$$E[X] = \sum_{i=1}^{\infty} x_i p(x_i)$$

- O que é melhor
  - Receber R$2,00 com uma probabilidade de 0,37; ou
    - E[x] = 2 * 0,37 = 0,74
  - Receber R$1,00 com uma probabilidade de 0,94 ?
    - E[x] = 1 * 0,94 = 0,94

Valor esperado de um bilhete de loteria

* O valor esperado total para o bilhete de USUS$1 do Illinois Dugout Doubler (arredondado para o centavo mais próximo) é o seguinte: $\frac{1}{15}$ (US$2) + $\frac{1}{42,86}$ (US$4) + $\frac{1}{75}$ (US$5) + $\frac{1}{200}$ (US$10) + $\frac{1}{300}$ (US$25) + $\frac{1}{1.589,40}$ (US$50) + $\frac{1}{8.000}$ (US$100) + $\frac{1}{16.000}$ (US$200) + $\frac{1}{48.000}$ (US$500) + $\frac{1}{40.000}$ (US$1.000) = US$0,13 + US$0,09 + US$0,07 + US$0,05 + US$0,08 + US$0,03 + US$0,01 + US$0,01 + US$0,01 + US$0,03 = US$0,51. No entanto, há também a chance de $\frac{1}{10}$ de ganhar um bilhete grátis, que tem um retorno esperado de US$0,51, de modo que o retorno esperado total é US$0,51 + $\frac{1}{10}$ (US$0,51) = US$0,51 + US$0,05 = US$0,56.

Aproximação de uma distribuição Binomial por uma Normal

- Aproximação da B(n, π) pela N(n π, n π(1 – π))

- Ou seja

$$\mu = n\pi \qquad\qquad \sigma = \sqrt{n\pi(1-\pi)}$$

- A aproximação da distribuição binomial pela normal é boa quando
  - n π(1 – π) ≥ 3

Valor esperado de uma Distribuição Binomial

Se $X \sim \mathrm{B}(n, p)$ podemos escrever $X = X_1 + \cdots + X_n$

$$\mathrm{E}(X) = \mathrm{E}(X_1) + \cdots + \mathrm{E}(X_n) = np.$$

$$\begin{aligned} \mathrm{E}(X) &= 0 \times P(X = 0) + 1 \times P(X = 1) \\ &= 0 \times (1 - p) + 1 \times p = p. \end{aligned}$$

Valor esperado de uma Distribuição Binomial

A variância de X é definida por Var(X) = E(X^2) - [E(X)]^2. Temos que

$$
\begin{aligned}
\mathrm{E}(X^2) &= 0^2 \times P(X=0) + 1^2 \times P(X=1) \\
&= 0 \times (1-p) + 1 \times p = p.
\end{aligned}
$$

Assim, $\mathrm{Var}(X) = p - p^2 = p(1-p)$ e portanto $\mathrm{DP}(X) = \sqrt{p(1-p)}$.

Aproximação de uma distribuição Binomial por uma Normal

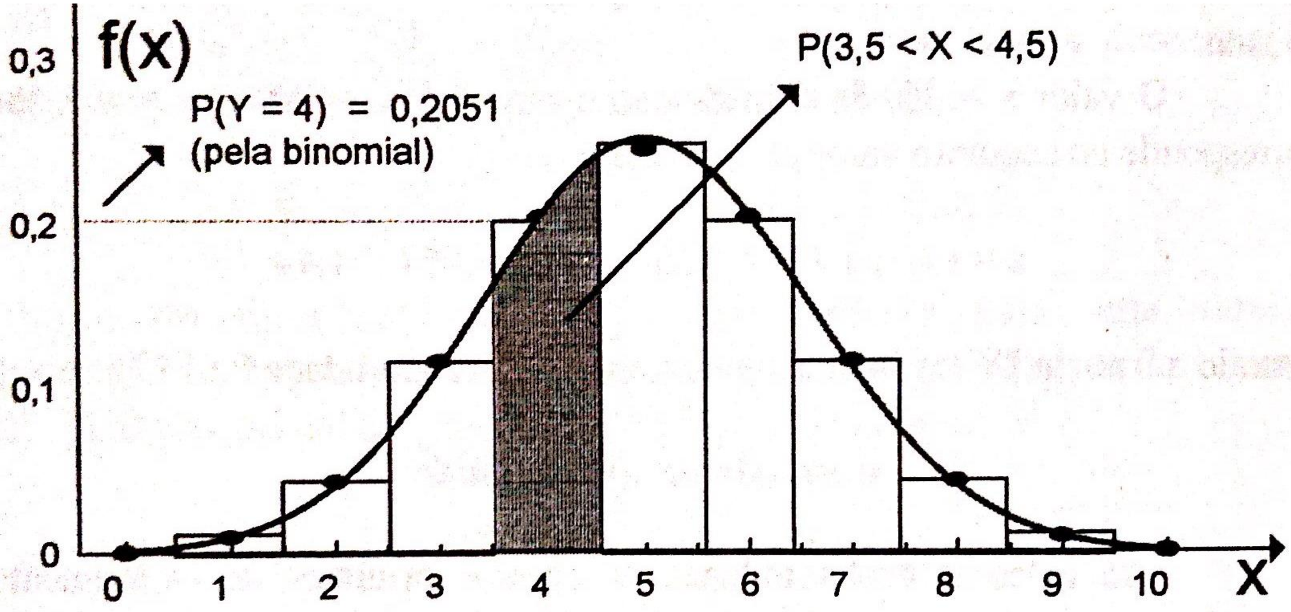
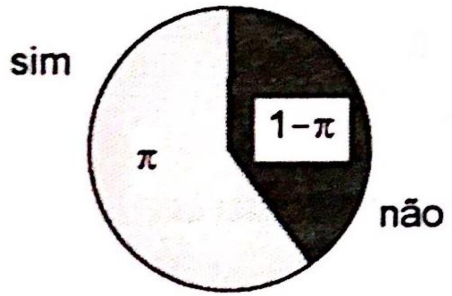


FIG. 8.11 Aproximação da probabilidade do evento Y = 4 (da distribuição binomial) para a probabilidade do evento 3,5 < X < 4,5 (da distribuição normal).

Usando adequadamente a distribuição normal, encontramos a probabilidade do evento $3{,}5 < X < 4{,}5$ como sendo igual a 0,2034. (Exercício: verifique o cálculo desta probabilidade).[5] Se fosse usada diretamente a distribuição binomial, chegaríamos à probabilidade igual a 0,2051 (Tabela II

Estimação de parâmetros

POPULAÇÃO	AMOSTRA
- todos os moradores do município; - os elementos da população estão divididos em *favoráveis* e *contrários* ao projeto; - parâmetro de interesse: π = *proporção de favoráveis.* 	- **n** moradores do município, selecionados aleatoriamente; - cada elemento da amostra é classificado como *favorável* ou *contrário* ao projeto; - estatística: **P** = *proporção de favoráveis na amostra*, isto é: $P = \dfrac{\text{nº de favoráveis na amostra}}{n}$

$$\text{Qual o valor de } \pi = ? \xleftarrow[\textit{processo de estimação}]{\hspace{3cm}} \pi = P \pm \textit{erro amostral}$$

Estimador Pontual

- O estimador pontual para **p (ou seja, π)**, também denominado proporção amostral, é definido como

$$\hat{p} = \frac{X}{n}$$

- X denota o número de elementos da amostra que apresentam a característica analisada
- n representa o tamanho da amostra coletada

Exemplo de Estimador Pontual

- Numa amostra de 400 indivíduos
  - 240 são a favor de um projeto municipal

$$\hat{p} = \frac{240}{400} = 0{,}6$$

- Como estimar o erro amostral ?
  - Coletando várias amostras de 400 indivíduos
    - E analisando a distribuição do valor estimado da proporção

Uma simulação

Para ilustrarmos a distribuição amostral de **P**, conforme a situação da Figura 9.3, podemos simular várias amostras de tamanho n = 400, segundo o modelo especificado. A simulação pode ser executada com o apoio de uma tabela de números aleatórios (Tabela I do apêndice). Cada número de um algarismo, observado na tabela, simula a observação de um elemento da população, da seguinte forma.[2]

- Quando o algarismo extraído da tabela de números aleatórios for um valor do conjunto {0,1,2,3,4,5,6}, que acontece com probabilidade 7/10, simula a observação de um indivíduo *favorável* ao projeto.

- Quando o algarismo extraído da tabela de números aleatórios for um valor do conjunto {7,8,9}, que acontece com probabilidade 3/10, simula a observação de um indivíduo *contrário* ao projeto.

Simulação do calculo do estimador de p

Valores simulados de P

0,70	0,67	0,69	0,70	0,74	0,71	0,68	0,69
0,69	0,71	0,71	0,68	0,71	0,71	0,70	0,70
0,72	0,71	0,68	0,69	0,69	0,66	0,69	0,69
0,67	0,72	0,73	0,71	0,70	0,67	0,70	0,71
0,70	0,72	0,71	0,68	0,70	0,72	0,70	0,72
0,73	0,67	0,71	0,76*	0,73	0,70	0,68	0,66
0,73	0,69	0,69	0,68	0,69	0,71	0,69	0,74
0,74	0,70	0,70	0,67	0,70	0,71	0,69	0,72
0,70	0,69	0,73	0,70	0,74	0,72	0,70	0,68
0,69	0,70	0,74	0,75	0,71	0,69	0,71	0,70
0,72	0,65*	0,69	0,69	0,70	0,71	0,66	0,71
0,66	0,65	0,68	0,69	0,69	0,68	0,71	0,71
0,72	0,68	0,68	0,73				

* Valor máximo e valor mínimo.

Histograma de freqüências

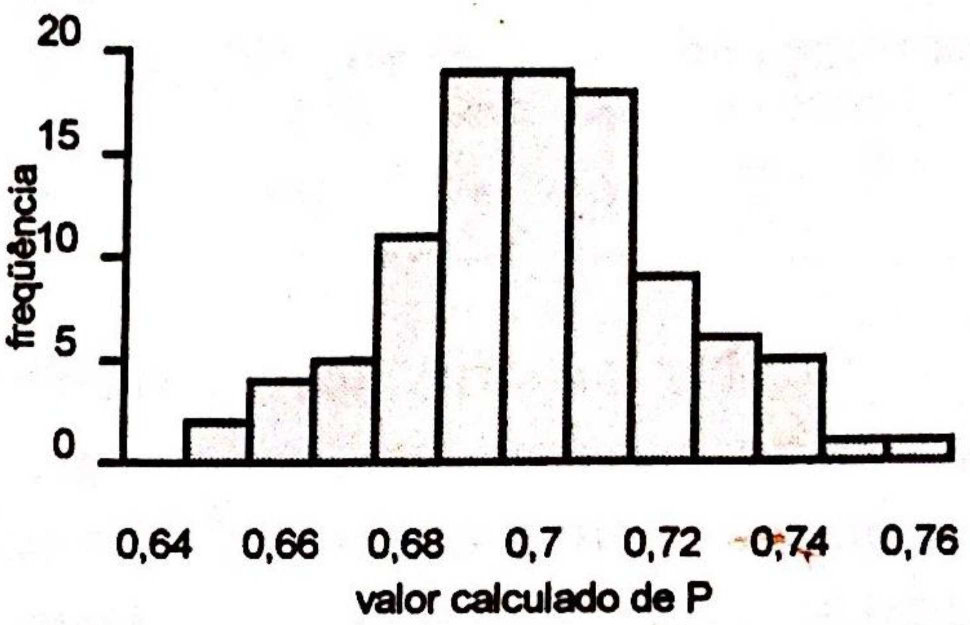


FIG. 9.4 Cem observações da distribuição amostral de **P**, considerando amostras de tamanho n = 400 e $\pi = 0{,}70$.

Simulação do calculo do estimador de p

- Nenhuma das amostras teve erro amostral
  - maior do que 0,6
- Então temos alta confiança de que
  - Raras serão as amostras com $\hat{p} > 0.76$ ou $\hat{p} < 0.64$
- O quão confiante estamos?
  - Podemos mensurar essa confiança ?
    - 68,2% dos valores estão dentro do intervalo $\mu \pm \sigma$
    - 95,4% dos valores estão dentro do intervalo $\mu \pm 2\sigma$
    - 99,7% dos valores estão dentro do intervalo $\mu \pm 3\sigma$
  - $\sigma = 0.02 \quad \mu = 0.7$

Simulação do calculo do estimador de p

- σ = 0.02   μ = 0.7
  - 68,2% dos valores de $\hat{p}$ estão dentro do intervalo (68 em 100)
    - 0.68 a 0.72
  - 95,4% dos valores de $\hat{p}$ estão dentro do intervalo (95 em 100)
    - 0.66 a 0.74
  - 99,7% dos valores de $\hat{p}$ estão dentro do intervalo (99 em 100)
    - 0.64 a 0.76

Cálculo do desvio padrão de do conjunto de amostras

- Considere a distribuição de amostras de P
  - Onde P = “proporção de uma característica da população”
- O desvio padrão dessa distribuição é chamado de
  - Erro padrão de P

$$S_p = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Premissas do cálculo do erro padrão

- Considera que seja válida a aproximação da distribuição binomial
  - Por uma distribuição normal
- Ou seja,
  - Desde que π não seja próximo de 0 (evento raro),
  - π não seja próximo de 1 (evento quase certo), e
  - $n \geq 30$
- A aproximação da distribuição binomial pela normal é boa quando
  - $n\,\pi(1 - \pi) \geq 3$

EXEMPLO 9.1, continuação - Admita que na amostra de n = 400 elementos, encontramos 60% de favoráveis. Temos, então, P = 0,60 (ou 60%) e erro padrão de P dado por

$$S_P = \sqrt{\frac{P.(1-P)}{n}} = \sqrt{\frac{(0,60).(0,40)}{400}} = 0,024.$$

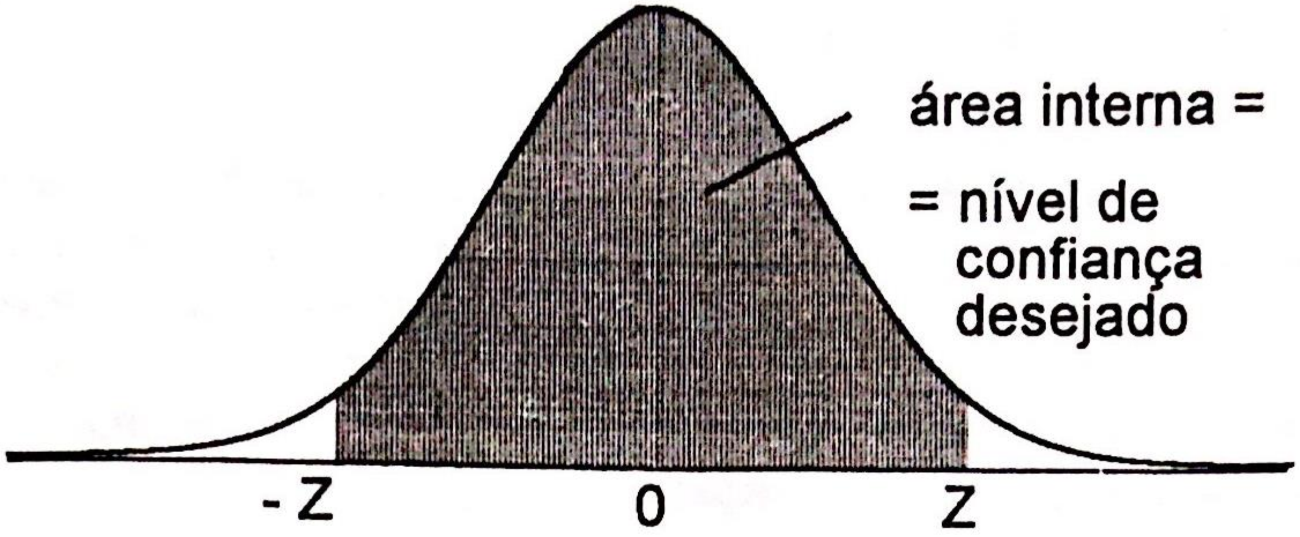
Usando nível de confiança de 95%, temos um erro amostral máximo provável de $2.S_P = 2.(0,024) = 0,048$ (ou 4,8%). Desta forma, podemos dizer que o intervalo: 60,0% ± 4,8% (isto é, o intervalo de 55,2% a 64,8%) contém, com 95% de confiança, o parâmetro π = *proporção de favoráveis em toda a população de moradores do município*.

O intervalo:

$$P \pm 2.S_P \qquad \text{(isto é, de } P - 2.S_P \text{ até } P + 2.S_P\text{)}$$

é dito um ***intervalo de 95% de confiança*** para o parâmetro π.

Outros intervalos de confiança



PARTE DE UMA TABELA NORMAL PADRÃO

Área	0,800	0,900	0,950	0,980	0,990	0.995	0,998
z	1,282	1,645	1,960	2,326	2,576	2,807	3,090

FIG. 9.7 Valores de **z** para alguns níveis de confiança.

Fixado o nível de confiança, podemos obter o correspondente valor de z, como ilustra a Fig. 9.7 e, a partir daí, calcular a estimativa do erro amostral máximo provável, **z.S_P**, e o intervalo de confiança para π:

$$\mathbf{P \pm z \,.\, S_P}$$

EXEMPLO 9.1, continuação - No exemplo em questão, poderíamos querer um nível de 99% de confiança. Então, pela tabela da Fig. 9.7, temos que área = 0,99 implica em z = 2,576, resultando no seguinte limite provável para o erro amostral: S_P = (2,576).(0,024) = 0,062 (ou 6,2%). Então, com 99% de confiança, o seguinte intervalo:

$$60{,}0\% \pm 6{,}2\%$$

Exercício de Avaliação – 9 (Para Casa)

4) Considerando o Exemplo 9.1, faça as seguintes modificações, executando, em cada caso, um intervalo de confiança para o parâmetro π. Discuta sobre a precisão das estimativas ao variar **n** e π.

a) nível de confiança de 90%, n = 400, com 60% de favoráveis na amostra.
b) nível de confiança de 95%, porém considerando que a amostra tenha sido de n = 1000 moradores, acusando 600 favoráveis.
c) nível de confiança de 95%, n = 400, com 80 favoráveis.
d) nível de confiança de 95%, n = 400, com 320 favoráveis.
e) nível de confiança de 95%, n = 400, com 200 favoráveis.