Aproximação da binomial pela normal

Objetivo

Verificar como a distribuição normal pode ser utilizada para calcular, *de forma aproximada*, probabilidades associadas a uma variável aleatória com distribuição binomial.

1. Introdução

Distribuição Binomial

- n ensaios Bernoulli independentes
- $P(S) = P(\text{Sucesso}) = p$

X : número de sucessos observados nos n ensaios

$\Rightarrow$ X tem distribuição binomial com parâmetros n e p

Notação: $X \sim b(n\,;\,p)$

Resultado: $X \sim b(n\,;\,p) \longrightarrow$

$E(X) = n\,p$

$Var\,(X) = n\,p\,(1 - p)$

Exemplo 1:

Uma moeda honesta é lançada n = 10 vezes em idênticas condições.

Determinar a probabilidade de ocorrer cara entre 40% e 70% das vezes, inclusive.

Seja X : número total de caras nos 10 lançamentos

"Sucesso" : ocorrência de cara

$p = P(S) = 0{,}5$ (moeda honesta)

$\Rightarrow$ $X \sim b(10\ ;\ 0{,}5)$

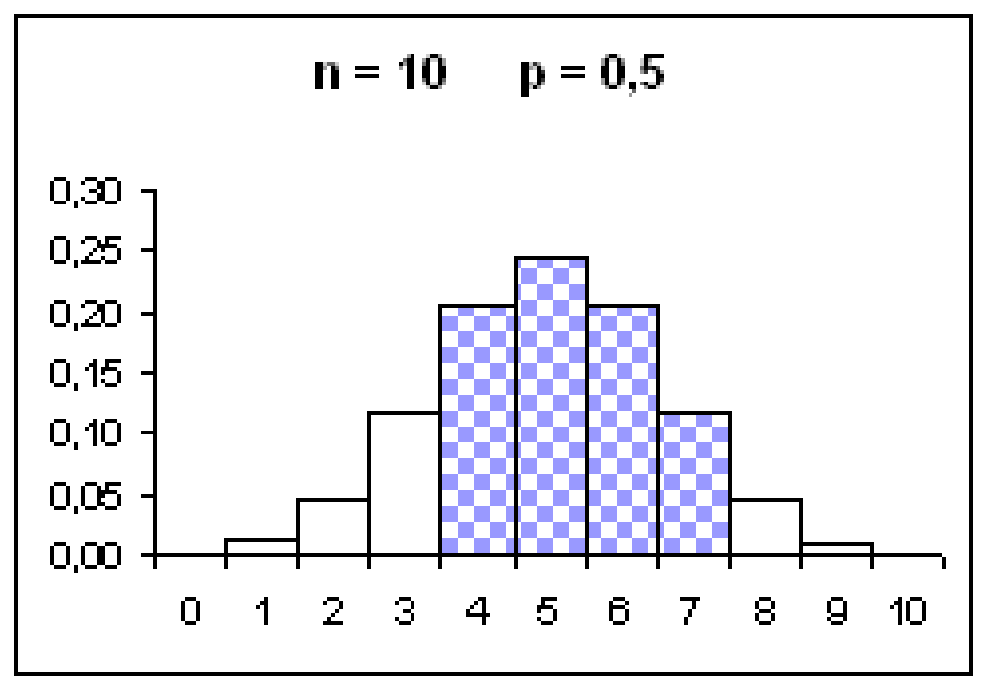
Probabilidade a ser calculada: $P(4 \leq X \leq 7)$

Distribuição de Probabilidades de $X \sim b(10 ; 0,5)$

Probability Density Function

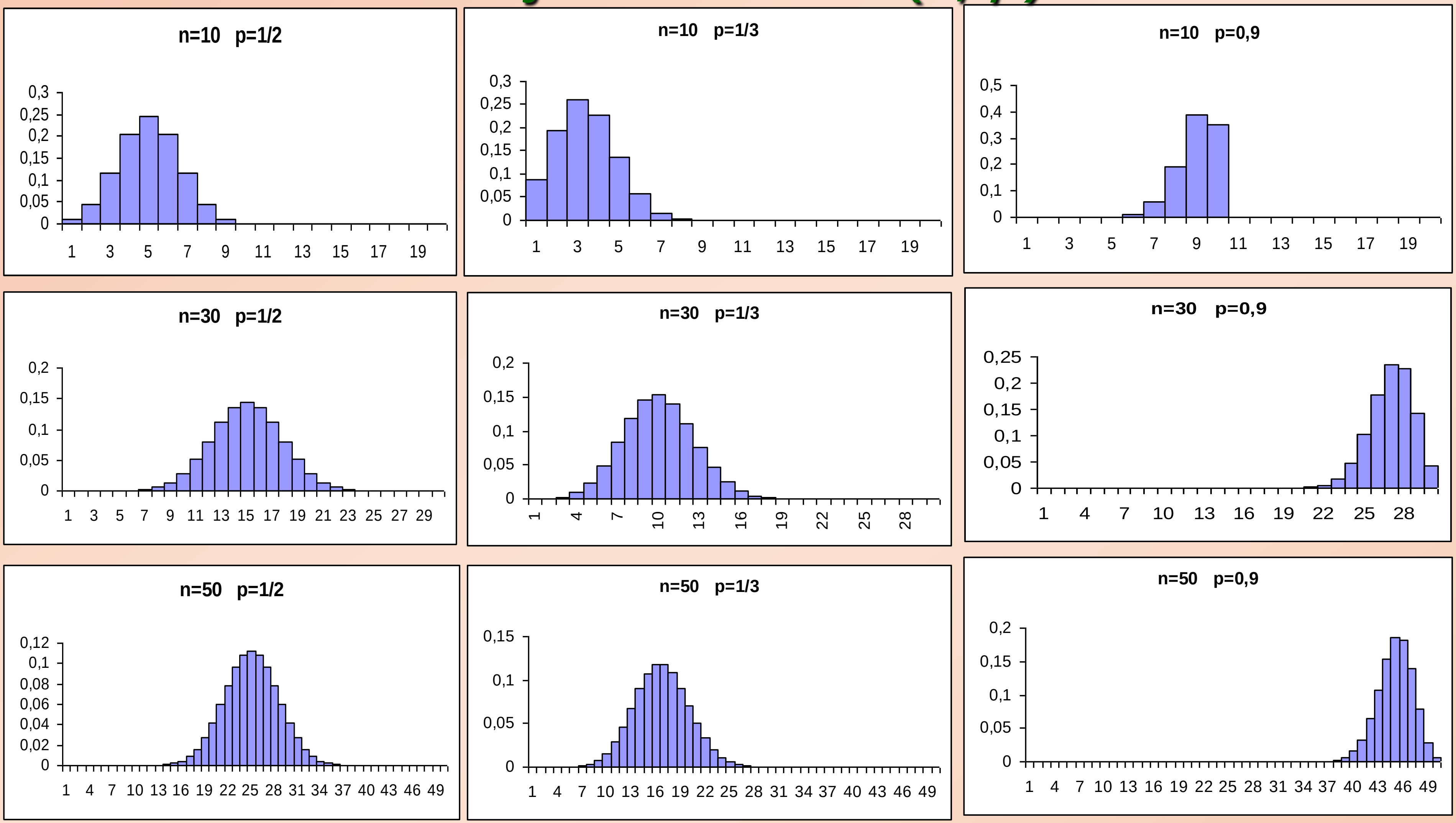
Binomial with n = 10 and p = 0,50

x	P($X = x$)
0	0,0010
1	0,0098
2	0,0439
3	0,1172
4	0,2051
5	0,2461
6	0,2051
7	0,1172
8	0,0439
9	0,0098
10	0,0010



$P(4 \leq X \leq 7) = 0,2051 + 0,2461 + 0,2051 + 0,1172 = 0,7735.$

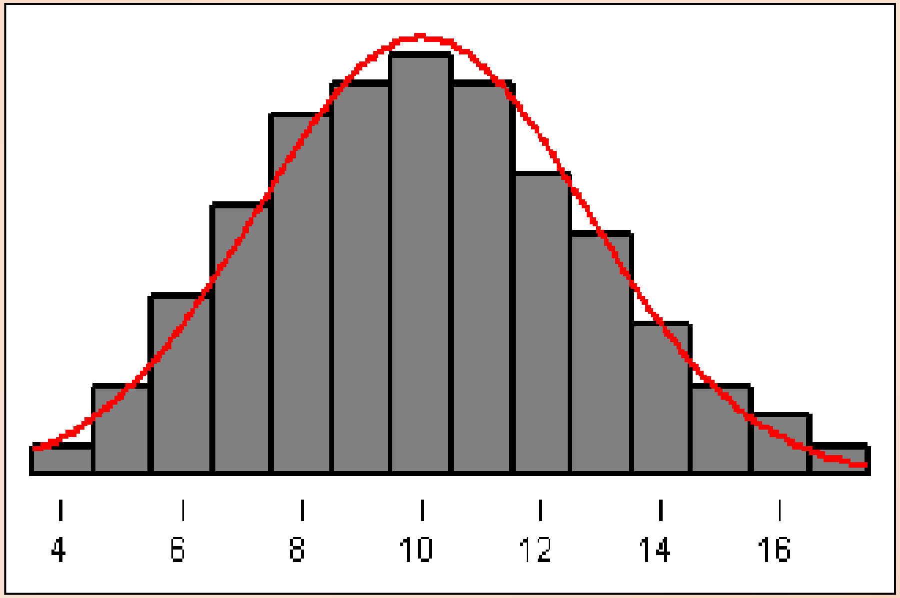
Distribuições binomiais (*n*, *p*)



Para *p* fixado, a medida que *n* cresce, os histogramas vão se tornando mais simétricos e com a forma da curva Normal.

2. Aproximação da binomial pela normal

Considere a binomial com n = 50 e p = 0,2, representada pelo histograma



P(Y=13) é igual a área do retângulo de base unitária e altura igual a P(Y=13); similarmente, P(Y=14), etc...
Logo, P($Y \geq$13) é igual à soma das áreas dos retângulos correspondentes.
A idéia é aproximar tal área pela área sob uma curva normal, à direita de 13. → ***Qual curva normal?***

$X \sim b(n\ ;\ p)$ $\Rightarrow$ $E(X) = np$
$Var(X) = np(1 - p)$

Parece razoável considerar a <u>*normal com média e variância iguais às da binomial*</u>, ou seja,
aproximamos a distribuição de probabilidades de *X* pela distribuição de probabilidades de uma variável aleatória *Y*, sendo

$$Y \sim N(\ \mu_y\ ;\ \sigma_y^2) \text{ com } \mu_y = np \text{ e } \sigma_y^2 = np(1 - p).$$

Portanto,

- $P(\ a \le X \le b) \approx P(a \le Y \le b)$
- $P(\ X \ge a) \approx P(Y \ge a)$
- $P(\ X \le b) \approx P(Y \le b)$

com $Y \sim N(np;\ np(1 - p)\ )$.

O cálculo da probabilidade aproximada é feito da forma usual para a distribuição normal:

$$P(a \leq X \leq b) \approx P(a \leq Y \leq b) \text{ com } Y \sim N(np\,;\, np(1-p)).$$

Lembrando que

$$Z = \frac{Y - np}{\sqrt{np(1-p)}} \sim N(0;1),$$

então

$$P(a \leq Y \leq b) = P\left(\frac{a - np}{\sqrt{np(1-p)}} \leq \frac{Y - np}{\sqrt{np(1-p)}} \leq \frac{b - np}{\sqrt{np(1-p)}}\right)$$

$$= P\left(\frac{a - np}{\sqrt{np(1-p)}} \leq Z \leq \frac{b - np}{\sqrt{np(1-p)}}\right).$$

Exemplo 2: $X \sim b(225\ ;\ 0{,}2)$ $n = 225$ e $p = 0{,}2$

$E(X) = np = 225 \times 0{,}2 = 45$

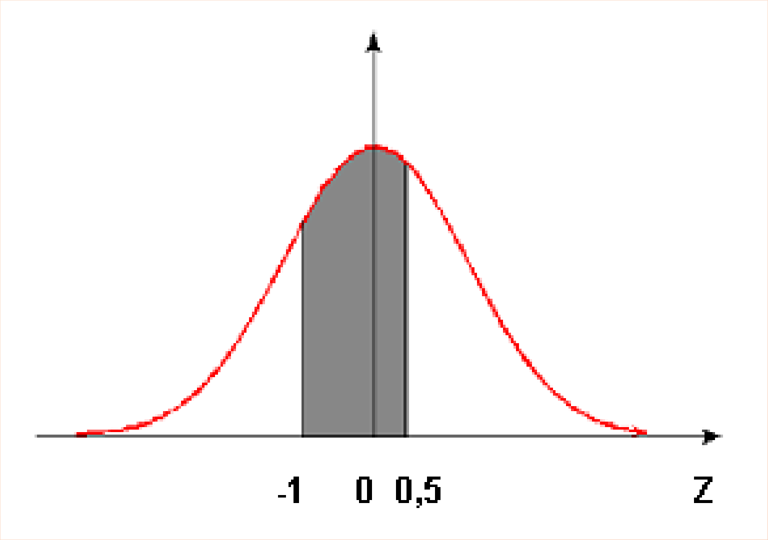
$Var(X) = np(1 - p) = 225 \times 0{,}2 \times 0{,}8 = 36$

$\Rightarrow$ $Y \sim N(45\ ;\ 36)$

a) $P(39 \leq X \leq 48) \approx P(39 \leq Y \leq 48)$

$$= P\left(\frac{39 - 45}{6} \leq \frac{Y - 45}{6} \leq \frac{48 - 45}{6}\right)$$

$$= P(-1{,}0 \leq Z \leq 0{,}5) = 0{,}5328$$

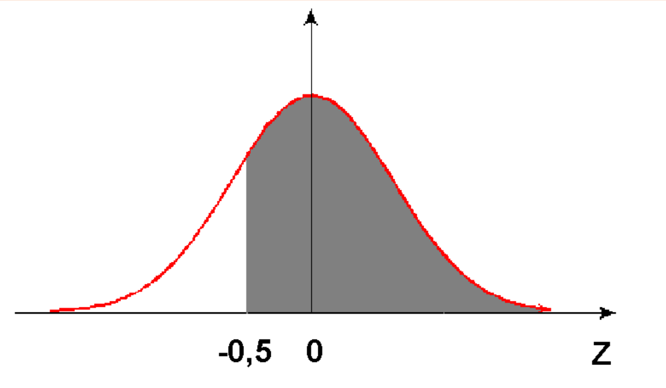


Probabilidade exata = 0,5853 (usando a distribuição binomial).

b) $P(X \geq 42) \approx P(Y \geq 42) = P\left(Z \geq \frac{42-45}{6}\right)$

$= P(Z \geq -0{,}5) = 0{,}6915$

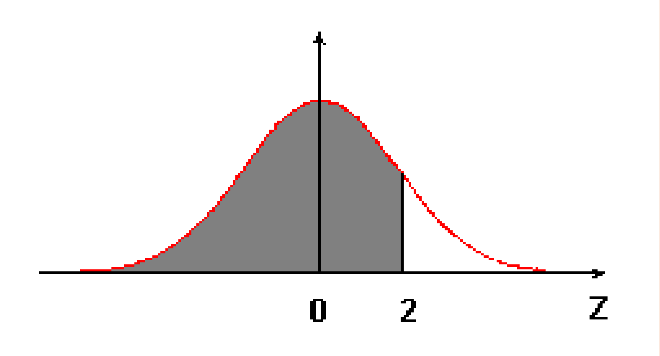
.



Probabilidade exata=0,7164(distr. binomial)

c) $P(X \leq 57) \approx P(Y \leq 57) = P\left(Z \geq \frac{57-45}{6}\right)$

$= P(Z \leq 2) = 0{,}9773.$

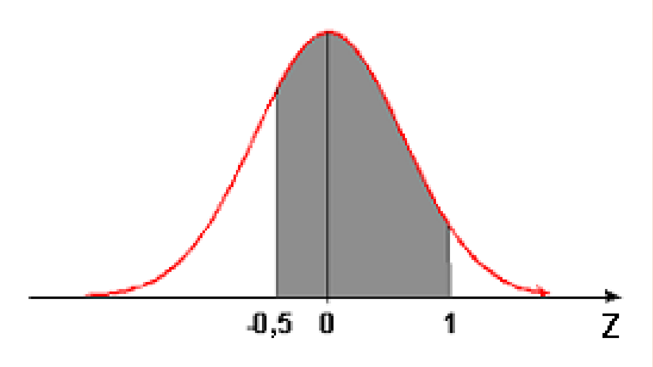


Probabilidade exata=0,9791(distr. binomial)

d) $P(41 < X < 52) = P(42 \leq X \leq 51) \approx P(42 \leq Y \leq 51)$

$= P(-0{,}5 \leq Z \leq 1)$

$= 0{,}5328.$



Probabilidade exata=0,5765(distr. binomial)

Observações :

1 - A aproximação da distribuição binomial pela normal é boa quando $np(1-p) \geq 3$.

2 - A demonstração da validade desta aproximação é feita utilizando-se o Teorema do Limite Central.

3 - A aproximação pode ser melhorada através do uso da "Correção de Continuidade".

Exemplo 3:

Um sistema é formado por 100 componentes, cada um dos quais com confiabilidade (probabilidade de funcionar adequadamente num certo período) igual a 0,9.

Se esses componentes funcionarem de forma independente um do outro e se o sistema funcionar adequadamente enquanto pelo menos 87 componentes estiverem funcionando, qual é a confiabilidade do sistema?

(Usar a aproximação normal)

X : número de componentes que funcionam adequadamente.

$X \sim \mathrm{b}(100;\, 0{,}9)$ $\Rightarrow$ $\mathrm{E}(X) = np = 100 \times 0{,}9 = 90$

$n = 100 \quad p = 0{,}9$ $\quad \mathrm{Var}(X) = np(1-p) = 100 \times 0{,}9 \times 0{,}1 = 9$

Confiabilidade do sistema: $\mathrm{P}(X \geq 87)$

$\mathrm{P}(X \geq 87) \approx \mathrm{P}(Y \geq 87)$, sendo $Y \sim \mathrm{N}(90\,;\, 9)$

$$\approx \mathrm{P}\left(\frac{Y-90}{3} \geq \frac{87-90}{3} \right) = \mathrm{P}\left(Z \geq -1 \right) = \mathrm{P}\left(Z \leq 1 \right)$$

$= 0{,}8413.$

Assim, a confiabilidade do sistema é aproximadamente igual a 0,8413.

Exemplo 4: Uma moeda honesta é lançada 100 vezes.

a) Calcular a probabilidade do número de caras estar entre 40% e 70% dos lançamentos, inclusive.

X : número de caras em 100 lançamentos $\Rightarrow$ $X \sim \text{b}(100\ ;\ 0{,}5\ )$

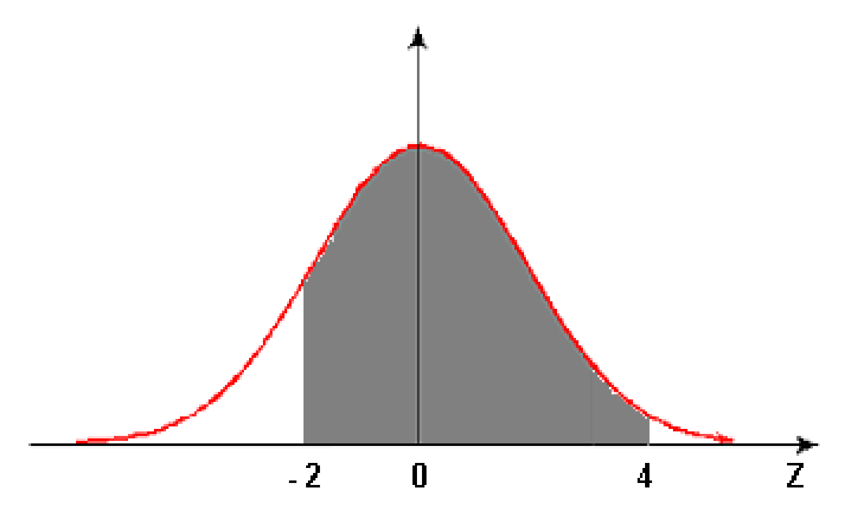
$\text{E}(X) = n\, p = 100 \times 0{,}5 = 50$ caras.

$\text{Var}(X) = n\, p\, (1 - p) = 100 \times 0{,}5 \times 0{,}5 = 25.$

$\text{P}(40 \leq X \leq 70\ ) \approx \text{P}(40 \leq Y \leq 70\ )$ (sendo $Y \sim \text{N}(50\ ;\ 25)$)

$$= \text{P}\left(\frac{40 - 50}{5} \leq \frac{Y - 50}{5} \leq \frac{70 - 50}{5}\right)$$

$$= \text{P}\left(-2 \leq Z \leq 4\right) = 0{,}9773.$$



Probabilidade exata= 0,9824.

b) Determinar um intervalo simétrico em torno do número médio de caras, tal que a probabilidade de observar um valor de *X* nesse intervalo é 80%.

Intervalo simétrico em torno da média: (50 – *a*, 50 + *a*)

$P(50 - a \leq X \leq 50 + a) = 0,8$

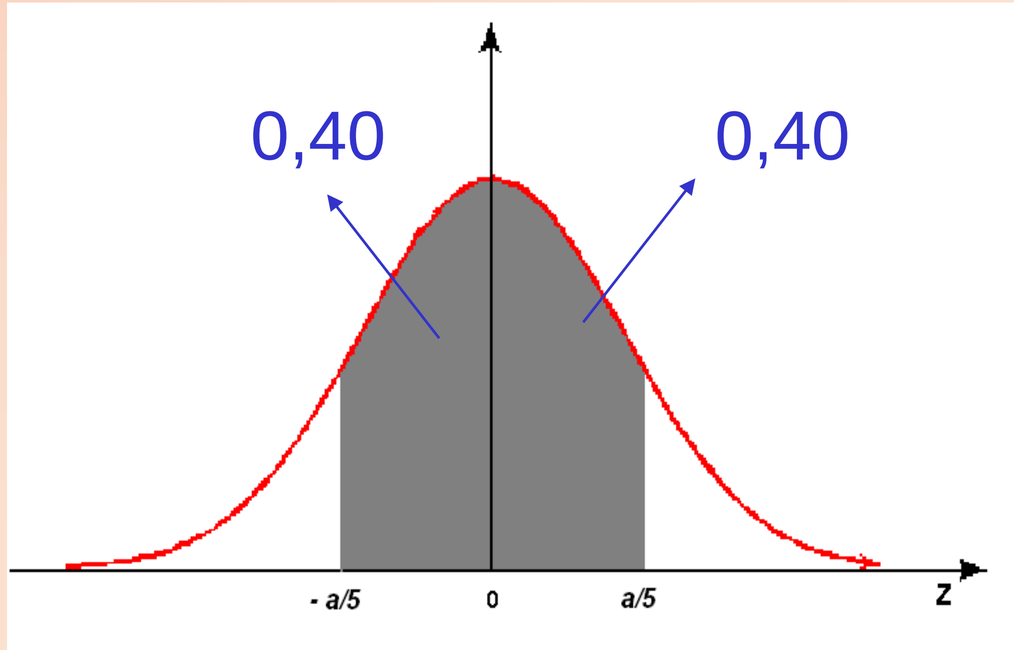
$P(50 - a \leq X \leq 50 + a) \approx P(50 - a \leq Y \leq 50 + a)$ $\quad Y \sim N(50\ ;\ 25)$

$$= P\left(\frac{50 - a - 50}{5} \leq \frac{Y - 50}{5} \leq \frac{50 + a - 50}{5}\right)$$

$$= P\left(\frac{-a}{5} \leq Z \leq \frac{a}{5}\right) = 0,8.$$

$a = ?$, tal que

$$P\left(\frac{-a}{5} \leq Z \leq \frac{a}{5}\right) = 0{,}8$$



$$\Rightarrow \frac{a}{5} = 1{,}28$$

$$\Rightarrow a = 6{,}4$$

Intervalo procurado: (50 - 6,4 ; 50 + 6,4) $\Rightarrow$ (43,6 ; 56,4).

A probabilidade de em 100 lançamentos termos entre 43 e 57 caras é aproximadamente 80%.

c) Um pesquisador, não conhecendo p = P(cara) , decide lançar a moeda 100 vezes e considerá-la desonesta se o número de caras for maior que 59 ou menor que 41.
Qual a probabilidade de considerar indevidamente a moeda como desonesta?

X : número de caras nos 100 lançamentos

$X \sim b(100 ; p)$, com p desconhecido para o pesquisador

P(considerar indevidamente a moeda como desonesta) =

$P(X > 59 \text{ ou } X < 41, \text{ quando } p = 0{,}5) =$

$P(X \geq 60 \text{ ou } X \leq 40, \text{ quando } p = 0{,}5) \approx P(Y \geq 60) + P(Y \leq 40),$

sendo $Y \sim N(50 ; 25)$

Esta probabilidade fica

$$P(Y \geq 60) + P(Y \leq 40) = P\left(\frac{Y - 50}{5} \geq \frac{60 - 50}{5}\right) + P\left(\frac{Y - 50}{5} \leq \frac{40 - 50}{5}\right)$$

$$= P(Z \geq 2) + P(Z \leq -2)$$

$$= 0{,}0455. \text{ (Interpretação??)}$$

Exemplo 5:

Uma prova é constituída de 20 testes com quatro alternativas cada. Um aluno não estudou a matéria e vai respondê-los ao acaso. Qual a probabilidade de acertar 50% ou mais das questões?

X : número de acertos

$X \sim b(20\ ;\ 0{,}25) \quad \Rightarrow \quad E(X) = np = 5$ e $Var(X) = np(1-p) = 3{,}75$

$P(X \geq 10) \approx P(Y \geq 10) \qquad Y \sim N(5\ ;\ 3{,}75)$

$$= P\left(\frac{Y - 5}{1{,}93} \geq \frac{10 - 5}{1{,}93}\right) = P(Z \geq 2{,}59) = 0{,}0048.$$

Repetir para 40 testes com quatro alternativas.

$X \sim b(40\ ;\ 0{,}25) \quad \Rightarrow \quad E(X) = n\,p = 10 \qquad Var(X) = n\,p\,(1-p) = 7{,}5$

$P(X \geq 20) \approx P(Y \geq 20) \qquad Y \sim N(10\ ;\ 7{,}5)$

$$= P\left(\frac{Y - 10}{2{,}75} \geq \frac{20 - 10}{2{,}75}\right) = P(Z \geq 3{,}63) = 0{,}0001.$$

Para 40 testes com cinco alternativas

$X \sim \text{b}(40\ ;\ 0{,}20) \quad \Rightarrow \quad \text{E}(X) = n\,p = 8$

$\text{Var}(X) = n\,p\,(1 - p) = 6{,}4$

$\text{P}(X \geq 20) \approx \text{P}(Y \geq 20) \qquad Y \sim \text{N}(8\ ;\ 6{,}4)$

$$= \text{P}\left(Z \geq \frac{20 - 8}{2{,}53}\right) = \text{P}(Z \geq 4{,}74) \approx 0{,}0000.$$