

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

⇒ X tem distribuição binomial com parâmetros n e p

Notação: $X \sim b(n ; p)$

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

$$E(X) = n p$$
$$\text{Var} (X) = n p (1 - p)$$

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

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

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

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

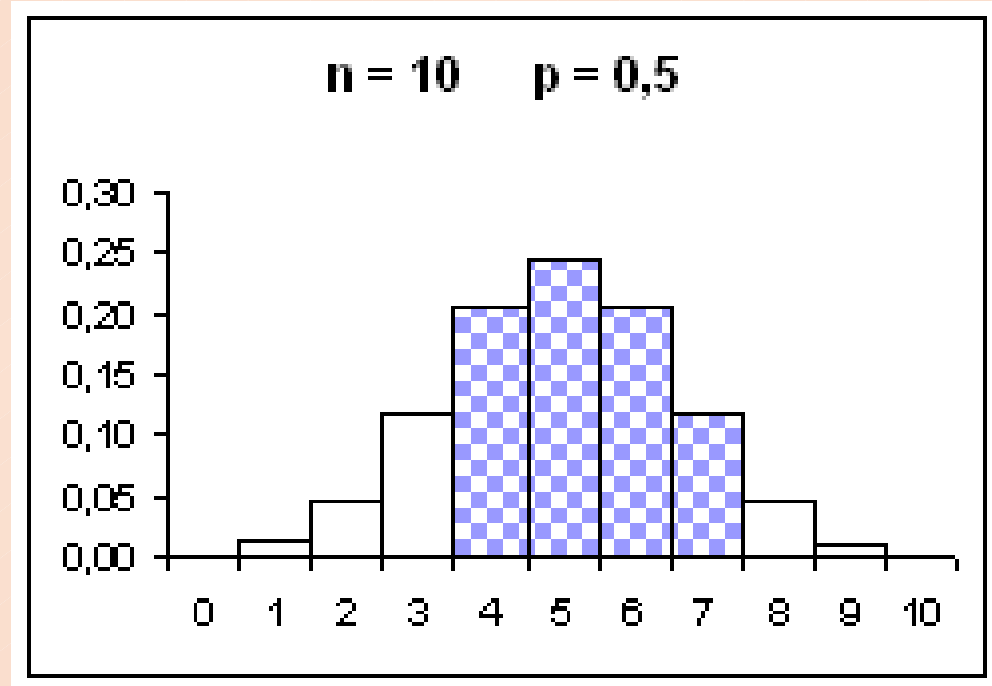
Distribuição de Probabilidades de $X \sim b(10 ; 0,5)$

MTB > pdf;

SUBC > bino 10 0,5.

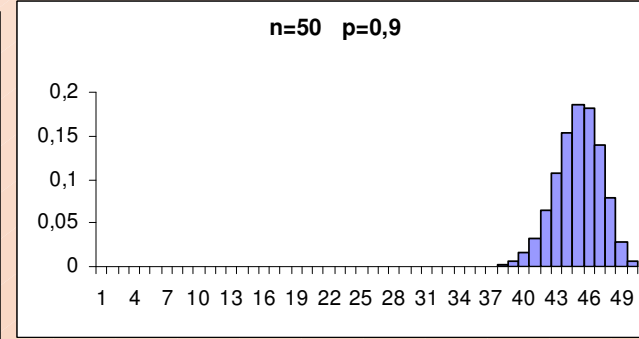
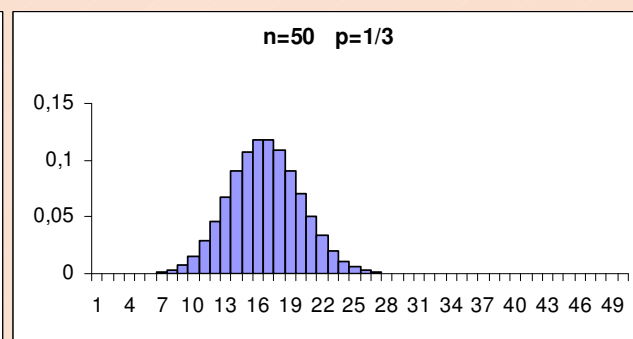
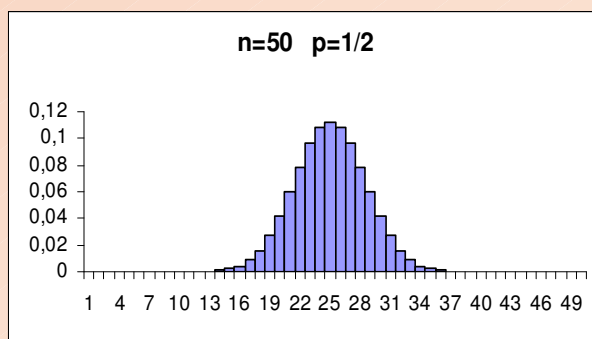
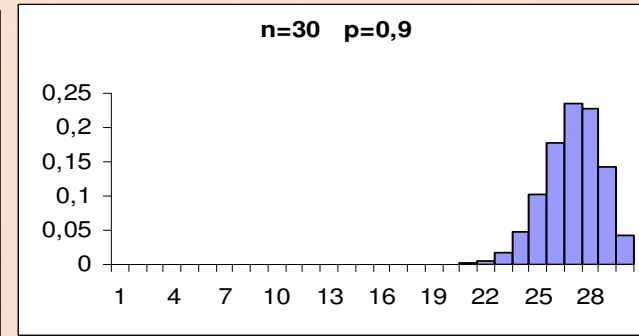
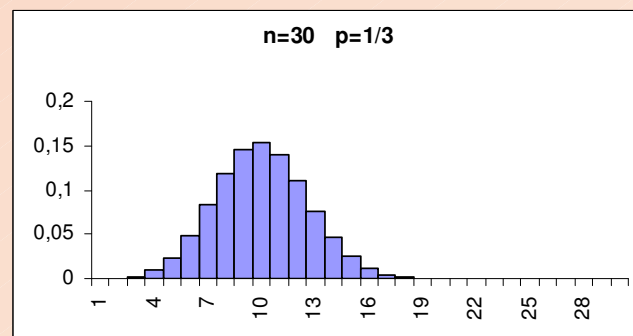
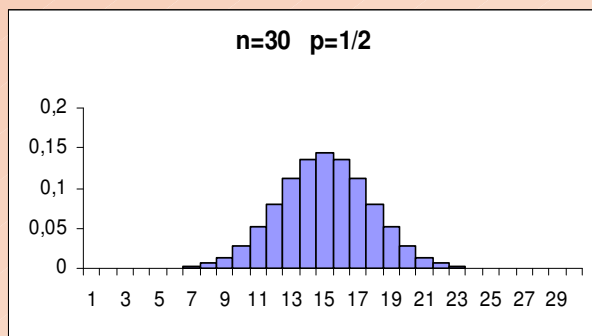
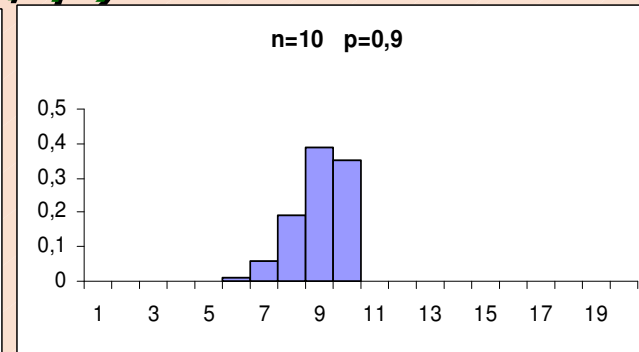
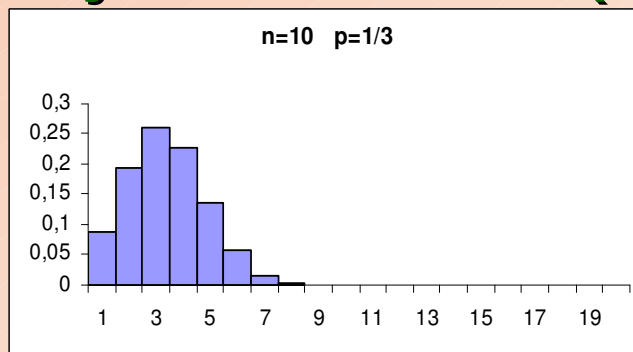
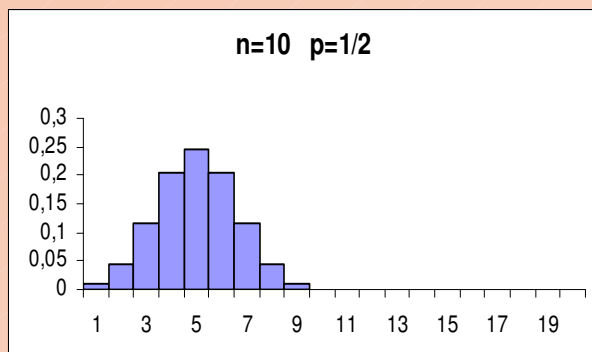
Probability Density Function
Binomial with $n = 10$ and $p = 0,50$

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



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

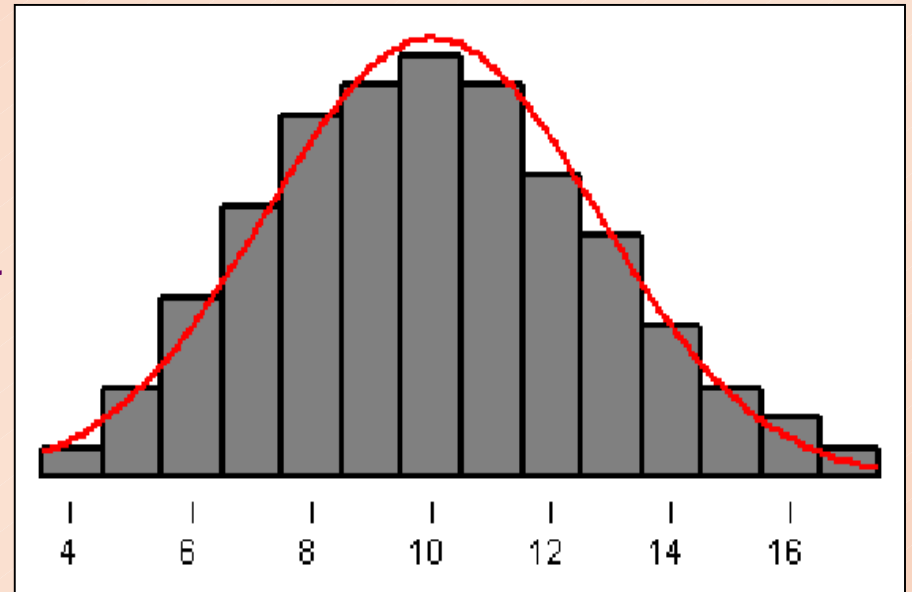
Distribuições binomiais (n, p)



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 $P(Y=13)$; semelhantemente, $P(Y=14)$, 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

$$E(X) = np$$

$$\text{Var}(X) = np(1 - p)$$

Parece razoável considerar a normal com média e variância iguais às da binomial, 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,

- $P(a \leq X \leq b) \approx P(a \leq Y \leq b)$
- $P(X \geq a) \approx P(Y \geq a)$
- $P(X \leq b) \approx P(Y \leq b)$

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

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

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

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

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

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

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

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

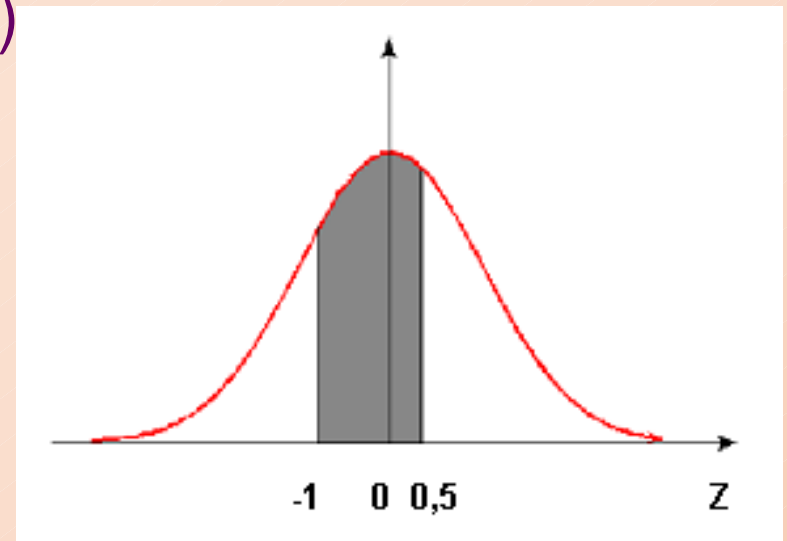
$$= P(-1,0 \leq Z \leq 0,5)$$

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

$$= P(Z \leq 0,5) - [1 - P(Z \leq 1,0)]$$

$$= A(0,5) - [1 - A(1,0)]$$

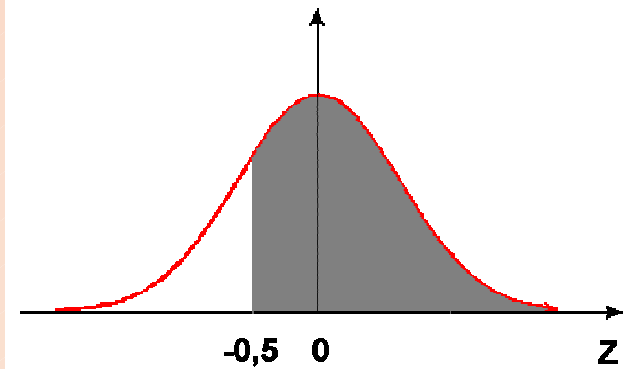
$$= 0,6915 - 0,1587 = 0,5328.$$



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

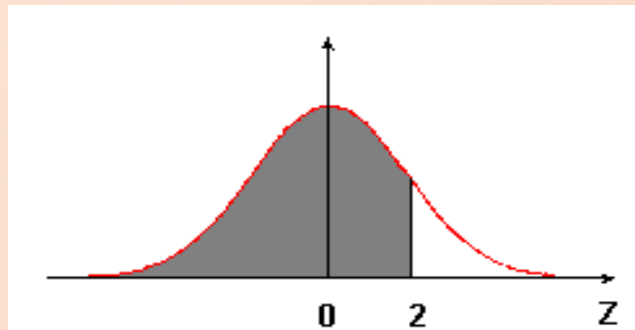
$$\begin{aligned}
 \text{b) } P(X \geq 42) &\approx P(Y \geq 42) = P\left(Z \geq \frac{42 - 45}{6}\right) \\
 &= P(Z \geq -0,5) = P(Z \leq 0,5) = A(0,5) \\
 &= 0,6915.
 \end{aligned}$$

Probabilidade exata=0,7164(distr. binomial)



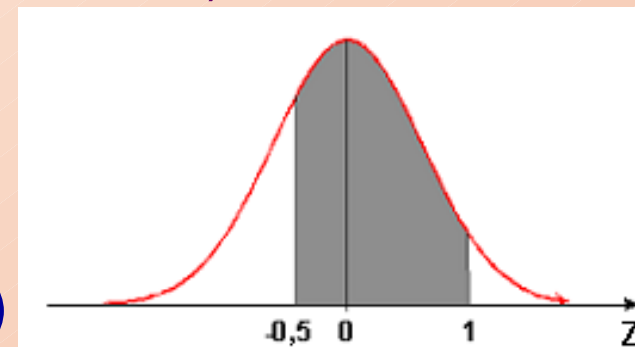
$$\begin{aligned}
 \text{c) } P(X \leq 57) &\approx P(Y \leq 57) = P\left(Z \geq \frac{57 - 45}{6}\right) \\
 &= P(Z \leq 2) = A(2) = 0,9773.
 \end{aligned}$$

Probabilidade exata=0,9791(distr. binomial)



$$\begin{aligned}
 \text{d) } P(41 < X < 52) &= P(42 \leq X \leq 51) \approx P(42 \leq Y \leq 51) \\
 &= P(-0,5 \leq Z \leq 1) \\
 &= A(1) - (1 - A(0,5)) = 0,5328.
 \end{aligned}$$

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.

$$X \sim b(100; 0,9)$$

$$E(X) = np = 100 \times 0,9 = 90$$

$$n = 100 \quad p = 0,9$$

\Rightarrow

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

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

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

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

$$= A(1) = 0,8413.$$

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 $\Rightarrow X \sim b(100 ; 0,5)$

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

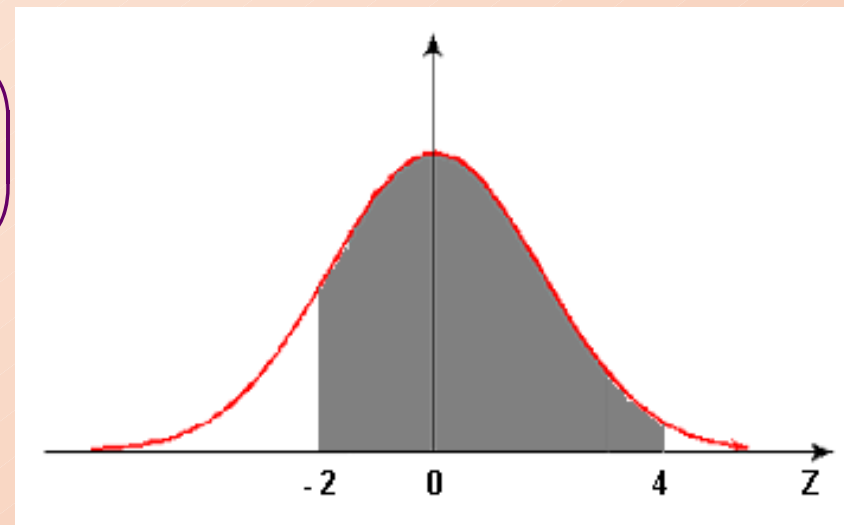
$Var(X) = n p (1 - p) = 100 \times 0,5 \times 0,5 = 25$.

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

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

$$= P(-2 \leq Z \leq 4) = 0,9773.$$

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: $(50 - a, 50 + a)$

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

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

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

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

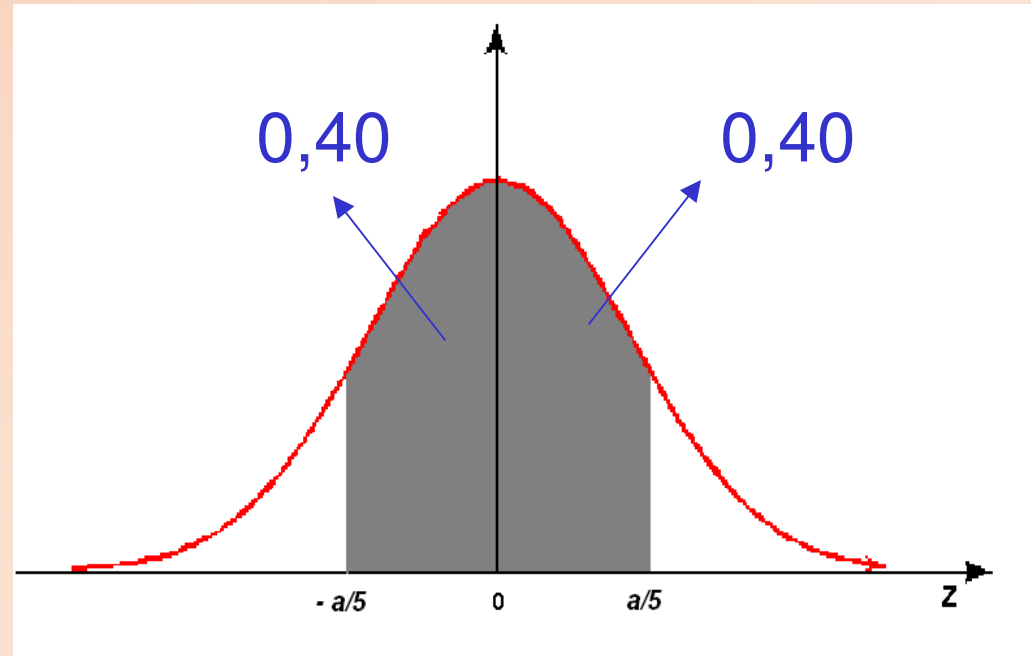
$a = ?$, tal que

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

$$\Rightarrow A\left(\frac{a}{5}\right) = 0,9$$

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

$$\Rightarrow a = 6,4$$



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 $p = P(\text{cara})$, 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

$P(\text{considerar indevidamente a moeda como desonesta}) =$

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

$P(X \geq 60 \text{ ou } X \leq 40, \text{ quando } p = 0,5) \approx P(Y \geq 60) + P(Y \leq 40)$,
sendo $Y \sim N(50; 25)$

Esta probabilidade fica

$$\begin{aligned} P(Y \geq 60) + P(Y \leq 40) &= P\left(\frac{Y - 50}{5} \geq \frac{60 - 50}{5}\right) + P\left(\frac{Y - 50}{5} \leq \frac{40 - 50}{5}\right) \\ &= P(Z \geq 2) + P(Z \leq -2) \\ &= 2(1 - A(2)) = 0,0455. \text{ (Interpretação??)} \end{aligned}$$

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

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

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

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

Repetir para 40 testes com quatro alternativas.

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

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

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

Para 40 testes com cinco alternativas

$$X \sim b(40 ; 0,20) \Rightarrow E(X) = n p = 8$$
$$\text{Var}(X) = n p (1 - p) = 6,4$$

$$P(X \geq 20) \approx P(Y \geq 20)$$

$$Y \sim N(8 ; 6,4)$$

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

Distribuição Normal : Valores de $P(Z \leq z) = A(z)$

	Segunda decimal de z									
	0	1	2	3	4	5	6	7	8	9
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000