Distribuições

Binomial e Hipergeométrica

Prof. A G de Mattos Neto

Distribuição Binomial

O Modelo da Caixa de Sapatos para o Cálculo de Probabilidades

Muitos problemas de probabilidade podem ser resolvidos por meio do “Modelo da Caixa de Sapatos”.

Suponha que você quer criar um modelo para o lançamento de uma moeda (justa), por 4 vezes. Esse modelo vai ser uma caixa de sapatos..., com tickets!

- Como você criaria esta caixa?
- Quantos tickets você colocaria na caixa?
- O que você escreveria nos tickets?
- Quantas retiradas você faria?
- Seriam retiradas com ou sem reposição?

O Modelo da Caixa

Tickets: 2, um “A” (cara) e outro “B” (coroa)
Retiradas: 4 vezes, com reposição

Quais seriam os resultados de suas retiradas? São 16 resultados possíveis. Como o número é pequeno, você faz por inspeção, ou construindo uma árvore de probabilidades!

4 caras:	AAAA					
3 caras:	BAAA	ABAA	AABA	AAAB		
2 caras:	BBAA	ABAB	BAAB	ABBA	BABA	AABB
1 cara:	ABBB	BABB	BBAB	BBBA		
0 caras:	BBBB					

O Modelo da Caixa

Mas..., como ajustaríamos essa caixa, se o objetivo fosse o de contar o número de caras em 4 lançamentos da moeda?

Ou, que variável aleatória (uma função do espaço amostral que leva cada resultado a um número) você usaria?

Tickets:	2, um "1" (cara) e outro "0" (coroa)
Retiradas:	4 vezes, com reposição

4 caras:	1111					
3 caras:	0111	1011	1101	1110		
2 caras:	0011	1010	0110	1001	0101	1100
1 cara:	1000	0100	0010	0001		
0 caras:	0000					

Sumarizando em uma tabela de chances

Nº de caras	0	1	2	3	4	Total
Nº possibilidades	1	4	6	4	1	16
Chance	1/16	4/16	6/16	4/16	1/16	1

Qual o valor esperado (o valor médio, ou a média da distribuição)?
Ou seja, quantas caras você espera tirar, em média, em 4 jogadas?

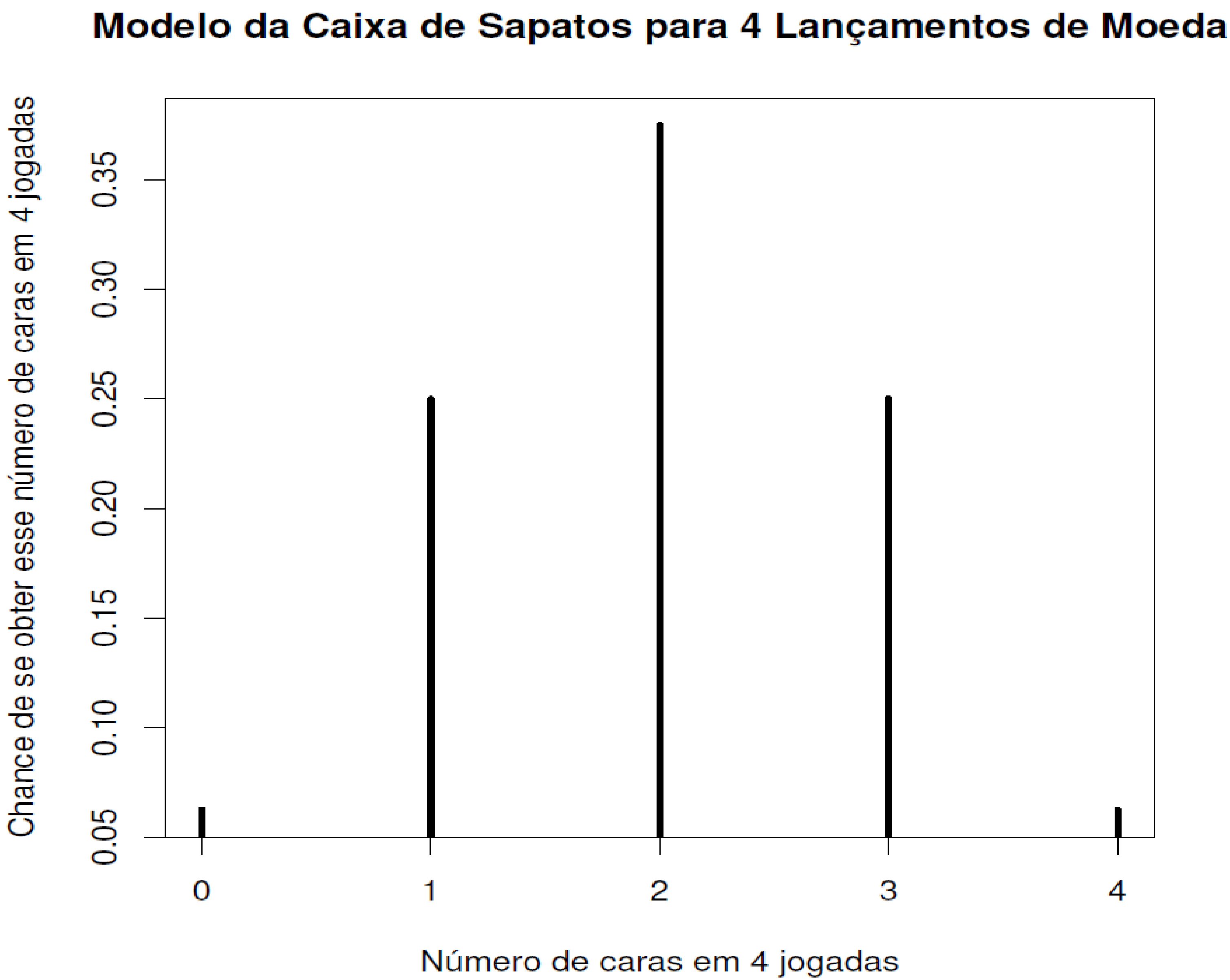
$(0x1 + 1x4 + 2x6 + 3x4 + 4x1) / 16 = 32 / 16 = 2$

$0x\text{Prob}(0) + 1x\text{Prob}(1) + 2x\text{Prob}(2) + 3x\text{Prob}(3) + 4x\text{Prob}(4) =$

$0x(1/16) \quad + 1x(4/16) \quad + 2x(6/16) \quad + 3x(4/16) \quad + 4x(1/16) \quad = 2$

E essa caixa fica assim, com a ajuda do R:

```
> x=c(0,1,2,3,4)
> px=c(1/16,4/16,6/16,4/16,1/16)
> x
[1] 0  1  2  3  4
> px
[1] 0.0625  0.2500  0.3725  0.2500  0.0625
> sum(px)
[1] 1
> plot(x,px)
> plot(x,px,type=“h”)
> plot(x,px,type=“h”,lwd=3,xlab=“Número de caras em 4 jogadas”,
        ylab=“Chance de se obter esse número de caras nessas 4
        jogadas”,main=“Modelo da Caixa de Sapatos para 4
        Lançamentos de Moeda”)
```


Modelo da Caixa de Sapatos para 4 Lançamentos de Moeda
Chance de se obter esse número de caras em 4 jogadas
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0
1
2
3
4
Número de caras em 4 jogadas

O Modelo da Caixa

Mas..., como ajustaríamos essa caixa, se o objetivo fosse o de contar o número de caras em 400 lançamentos da moeda?

Ou, que variável aleatória (uma função do espaço amostral que leva cada resultado a um número) você usaria?

- Como você criaria esta caixa?
- Quantos tickets você colocaria na caixa?
- O que você escreveria nos tickets?
- Quantas retiradas você faria?
- Seriam retiradas com ou sem reposição?
- Montar os resultados possíveis fica difícil?
- Como você resolveria isto?

Caixa com 400 jogadas da moeda, com a ajuda do R:

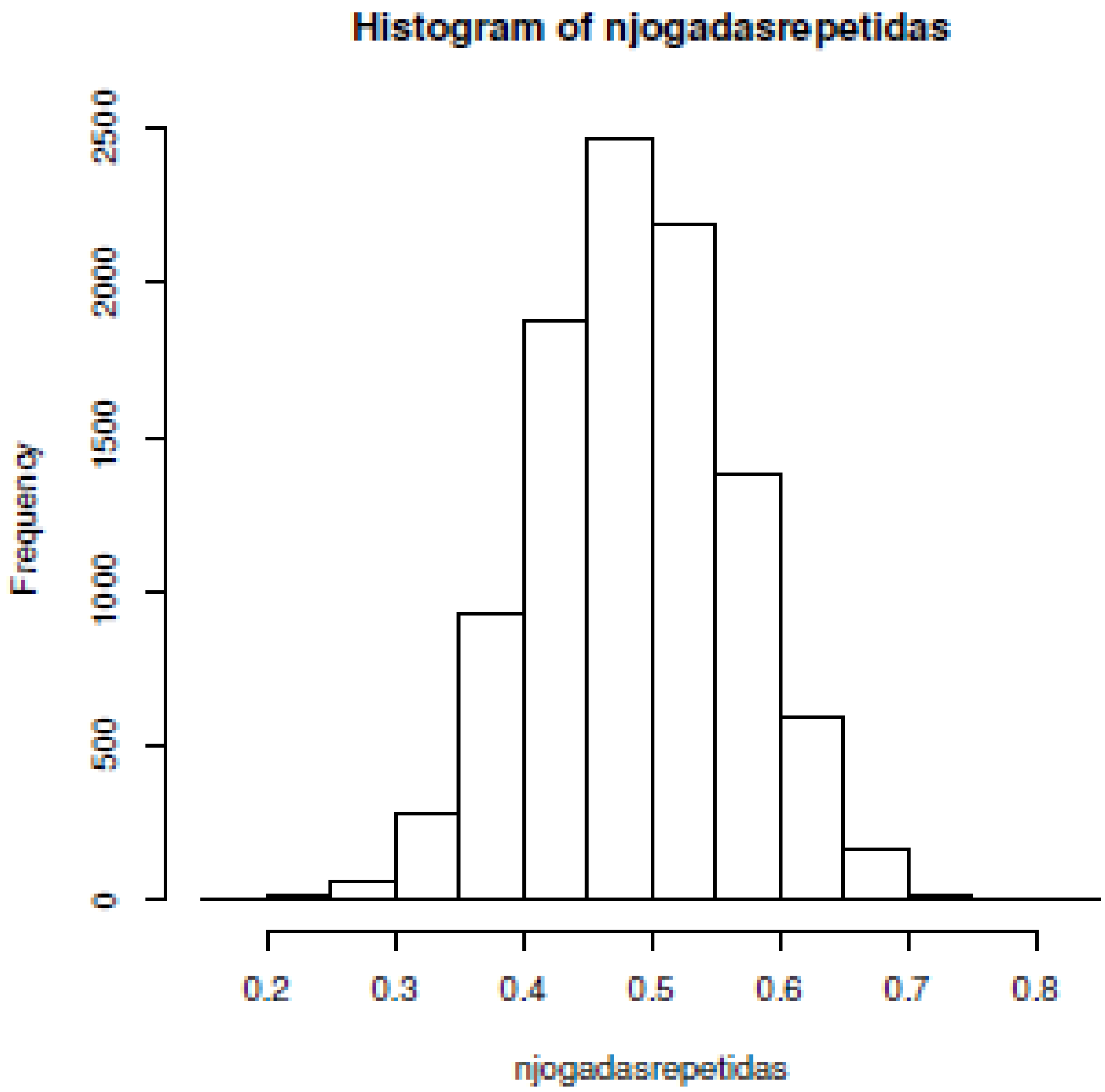
```
> moeda=c(0,1)
> voujogar4vezes=sample(moeda,4,replace=TRUE)
> voujogar4vezes
[1] 1  1  1  0
> voujogar4vezes=sample(moeda,4,replace=TRUE)
> voujogar4vezes
[1] 0  0  0  0
> voujogarnvezes=sample(moeda,40,replace=TRUE)
> voujogarnvezes
[1] 0  1  1  1  1  0  1  1  1  0  1  0  1  0  0  0  1  1  0  1  0  1  1  0  0  1  0
[28] 1  0  0   0  0  1  0  1  1  0  1  1  0
> sum(voujogarnvezes)
[1] 21
```

Caixa com 400 jogadas da moeda, com a ajuda do R:

```
> sum(voujogarnvezes)/40
[1] 0.525
> voujogarnvezes=sample(moeda,400,replace=TRUE)
> sum(voujogarnvezes)
[1] 203
> sum(voujogarnvezes)/400
[1] 0.5075
```

Como repetir este processo para obter a pdf (função massa de probabilidade) dessa caixa (experimento)?

```
> njogadasrepetidas=replicate(10000, sum (sample(moeda,40,
          replace=TRUE))/40)
> hist(njogadasrepetidas)
```



Histogram of njogadasrepetidas
Frequency
0
500
1000
1500
2000
2500
0.2
0.3
0.4
0.5
0.6
0.7
0.8
njogadasrepetidas

Voltando para a caixa no caso de 4 lançamentos de uma moeda

Nº de caras	0	1	2	3	4	Total
Nº possibilidades	1	4	6	4	1	16
Chance	1/16	4/16	6/16	4/16	1/16	1

Qual o valor esperado (o valor médio, ou a média da distribuição)?
Ou seja, quantas caras você espera tirar, em média, em 4 jogadas?

$(0x1 + 1x4 + 2x6 + 3x4 + 4x1) / 16 = 32 / 16 = 2$

$0x\text{Prob}(0) + 1x\text{Prob}(1) + 2x\text{Prob}(2) + 3x\text{Prob}(3) + 4x\text{Prob}(4) =$

$0x(1/16) \quad + 1x(4/16) \quad + 2x(6/16) \quad + 3x(4/16) \quad + 4x(1/16) \quad = 2$

Distribuições Binomial: 4 Lançamentos da Moeda

Para cada lançamento de uma moeda:

sucesso se for cara "1"	p
fracasso de for coroa "0"	$1 – p$

Por exemplo, 4 lançamentos resultando em 3 caras:

1 1 1 0 $\qquad P(1110) = p.p.p.(1-p)$

E como, para obter 3 caras em 4 lançamentos, há 4 possibilidades:

$$P(3 \text{ caras}) = P(1110) + P(1101) + P(1011) + P(0111)$$

Distribuições de Binomial: 4 Lançamentos da Moeda

$P(3 \text{ caras}) = p.p.p.(1-p) + p.p.(1-p).p + p.(1-p).p.p + (1-p).p.p.p$

$P(3 \text{ caras}) = 4\, p^3.(1-p)$

.... e assim por diante, tal que temos a seguinte distribuição para esses 4 lançamentos:

$P(0 \text{ caras}) = 1.(1-p)^4$
$P(1 \text{ cara}) = 4\, p.(1-p)^3$
$P(2 \text{ caras}) = 6\, p^2.(1-p)^2$
$P(3 \text{ caras}) = 4\, p^3.(1-p)$
$P(4 \text{ caras}) = 1\, p^4$

Você pode verificar que a soma dessas probabilidades totaliza 1.

Distribuição Binomial

$$n = número\ de\ tentativas$$
$$x = número\ de\ sucessos\ nas\ n\ tentativas$$
$$p = probabilidade\ de\ sucesso\ em\ cada\ tentativa$$

Função massa de probabilidade

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

Valor esperado (média) e variância

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

$$\sigma^2 = V(X) = np(1-p)$$

Distribuição Binomial no R

n = número de tentativas
x = número de sucessos obtidos nessas n tentativas
p = probabilidade de se obter sucesso em cada uma tentativa

dbinom(x,n,p)	densidade de probabilidade dado x = q
pbinom(x,n,p)	probabilidade cumulativa dado x = q
qbinom(p,n,p)	quantil q dado o percentil p
rbinom(r,n,p)	geração de r números randômicos

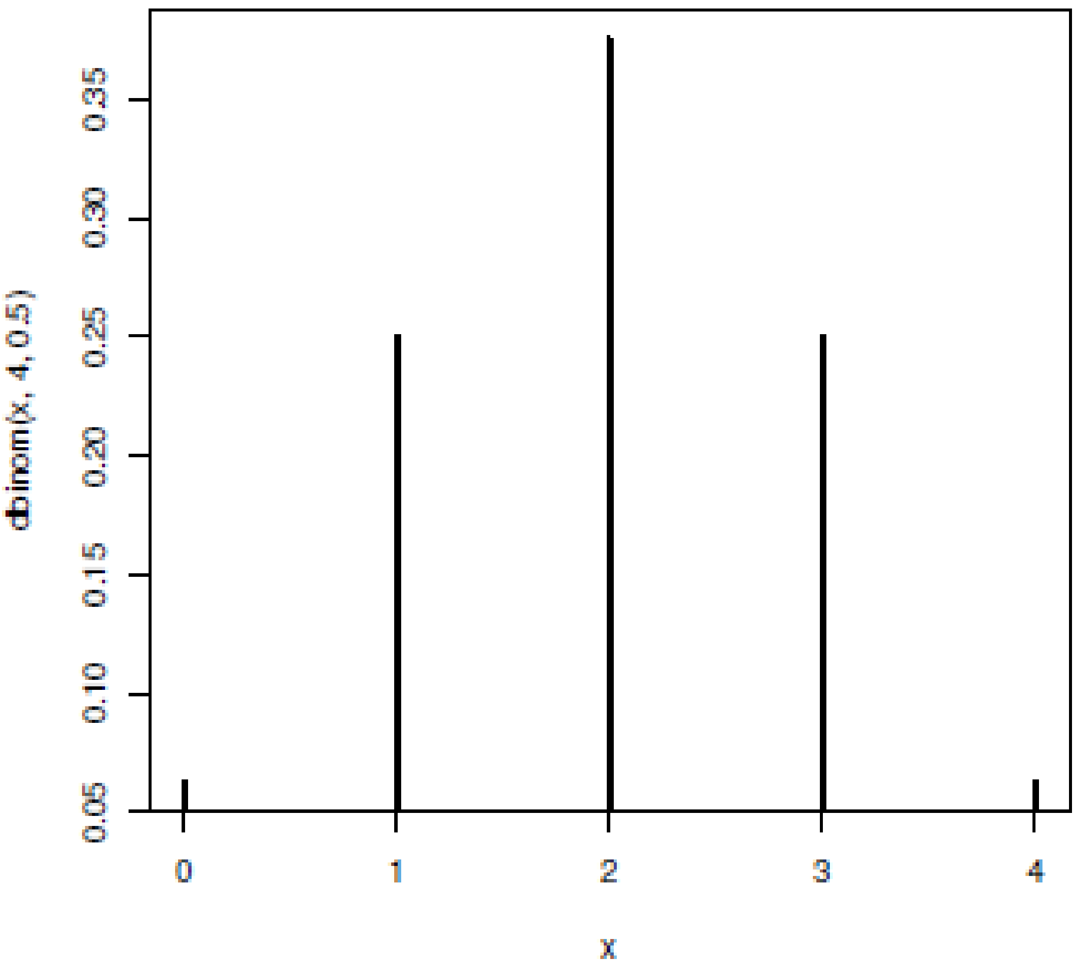
A caixa com 4 ou n lançamentos da moeda

Novamente com o R, e agora com a distribuição binomial!

```
> p0=dbinom(0,4,0.5)
> p1=dbinom(1,4,0.5)
> p2=dbinom(2,4,0.5)
> p3=dbinom(3,4,0.5)
> p4=dbinom(4,4,0.5)
> px=c(p0,p1,p2,p3,p4)
> px
[1] 0.0625  0.2500  0.3725  0.2500  0.0625
```

Para obter o plot dessa distribuição:

```
>x=c(0,1,2,3,4)
>plot(x,dbinom(x,4,0.5),type="h",lwd=3)
```



dbinom(x, 4, 0.5)
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0
1
2
3
4
x

Distribuição Hipergeométrica

Vinte-Um ou Black Jack

Se joga com 1 a 8 baralhos. Nos cassinos americanos, 7 baralhos.

Cada baralho tem 52 cartas, resultando em 7 x 52 = 364 cartas.

Vamos ver qual a chance de se tirar duas cartas de copas em duas retiradas sucessivas.

Como você montaria sua caixa? Com ou sem reposição?

Favor trabalhe em grupo.

Tempo: 3 minutos!

Vinte-Um ou Black Jack

91 tickets [1]
273 tickets [0]

Caixa de Sapatos

13 cartas de copas em cada baralho x 7 baralhos = 91
Então os outros naipes dão um total de 364 – 91 = 273

Nesse caso, vocês podem simplificar a caixa para esta aqui?

1 ticket [1]
3 tickets [0]

Caixa de Sapatos

Vinte-Um ou Black Jack

P(0) = probabilidade de tirar 0 cartas de copas em duas retiradas
P(1) = probabilidade de tirar 1 carta de copas em duas retiradas
P(2) = probabilidade de tirar 2 cartas de copas em duas retiradas

$P(x) = (p1^{a}vez).(p2^{a}vez)$

$P(0) = P([0][0]) = (273/364).(272/363)$

$P(1) = P([0][1])+P([1][0]) = (273/364).(91/363)+(91/364).(273/363)$

$P(2) = P([1][1]) = (91/364).(90/363)$

$P(0) + P(1) + P(2) = 1$

Distribuição Hipergeométrica

$$a = \textit{número de objetos considerados positivos}$$
$$b = \textit{número de objetos considerados negativos}$$
$$n = \textit{número de retiradas (tentativas sem reposição)}$$
$$N = \textit{número total de objetos no experimento}$$
$$x = \textit{número de sucessos nas } n \textit{ retiradas}$$

Função massa de probabilidade

$$f(x) = \frac{\binom{a}{x}\binom{b}{n-x}}{\binom{N}{n}}$$

Valor esperado (média) e variância

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

$$\sigma^2 = V(X) = np(1-p)\left(\frac{N-n}{N-1}\right)$$

$$onde\ p = a/N$$

Vinte-Um ou Black Jack

Usando a fórmula da distribuição hipergeométrica:

- número de objetos $N = 364$
- número de objetos positivos $a = 91$
- número de tentativas $n = 2$
- número de sucessos $x = 2$

1º termo numerador = $a!/x!(a-x)!=91!/2!89! = 91.90/2$

2º termo numerador = $b!/(n-x)!/(b-(n-x))!=273!/0!273!=1$

denominador = $N!/n!(N-n)!=364!/2!362!=364.363/2$

$P(2) = (91.90/2) \times 1 / (364.363/2) = (91/364).(90/363)$ ... Ok! Ufa!

Distribuição Hipergeométrica no R

a = número de objetos positivos [1]
b = número de objetos negativos [0]
n = número de retiradas
x = número de sucessos obtidos nessas k retiradas

dhyper(x,a,b,n)	densidade de probabilidade dado x = q
phyper(x,a,b,n)	probabilidade cumulativa dado x = q
qhyper(x,a,b,n)	quantil q dado o percentil p
rhyper(r,a,b,n)	geração de r números randômicos

Distribuição Binomial - Exemplos

dbinom(x,n,p)

x = número obtido de sucessos
n = número de tentativas ou observações
p = probabilidade de obter sucesso em cada tentativa ou observação

1. Jogue uma moeda 10 vezes. Seja x o número de caras obtidas nessas 10 jogadas.

2. Um tear produz 1% de itens defeituosos. Seja x o número de itens defeituosos nos próximos 25 itens produzidos.

3. Cada amostra de ar tem 10% de probabilidade de conter traços de uma determinada substância tóxica. Seja x o número de amostras que contém traços dessa substância nas próximas 18 amostragens.

Distribuição Binomial - Exemplos

4. De todos bits transmitidos em um canal de transmissão, dois em cada 1000 bits são transmitidos com erro. Seja x o número de bits transmitidos com erros nos próximos 1.000.000 de bits.

5. Um teste de múltipla escolha tem 10 questões, cada uma com quatro escolhas. Você tenta adivinhar cada questão. Seja x o número de questões que você acertou.

6. Nos próximos 20 nascimentos em um hospital, seja x o número de nascimento de meninas.

7. De todos pacientes sofrendo de uma determinada doença, 35% tiveram uma melhora significativa como resultado da administração de uma medicação. Nos próximos 100 pacientes administrados com esta medicação, seja x o número de pacientes que irão experimentar essa melhora significativa.

Distribuição Hipergeométrica - Exemplos

dhyper(x,a,b,n)

x = número obtido de sucessos
a = número de objetos (casos) positivos
b = número de objetos (casos) negativos
n = número de retiradas (tentativas sem reposição) ou observações

1. Uma urna contém 10 bolas verdes, 12 amarelas e 18 brancas. Seja x o número de bolas verdes obtidas de 10 retiradas consecutivas da urna (sem reposição).

2. Uma população de uma pequena cidade (com 5000 habitantes) mostra uma clara tendência de alta pressão arterial, com índice de 50 casos positivos para cada 100 habitantes. Seja x o número de habitantes com pressão alta nas próximas 800 observações dessa população.

Distribuição Hipergeométrica - Exemplos

3. Estima-se que 10% dos jovens (menores de 18 anos) de uma determinada cidade com 3000 jovens apresentará algum sintoma inicial de doença degenerativa antes de completar 45 anos. Seja x o número de jovens que se enquadrarão nesse caso dentre 100 selecionados para ume estudo de longo prazo sobre essa questão.

4. Maças são qualificadas através de teste de consistência antes de encaixotadas e despachadas para os grandes distribuidores, por meio de uma sonda com indicação da resistência à perfuração. Seja x o número de maças que seriam rejeitadas entre 50 selecionadas de um lote de 10000 unidades de determinado produtor, que consistentemente apresenta um índice de rejeição máximo de 1 maça em cada 500.

5. Um dia de produção de 850 peças fabricadas contém 50 peças que não satisfazem os critérios de qualidade de aceitação dos consumidores. Seja x o número dessas peças em uma seleção de 50 peças desse dia de produção.

Distribuição Hipergeométrica - Exemplos

6. Uma firma de auditoria contratada por um governo municipal desconfia que um auditor fiscal, o qual fiscalizou 470 empresas em um determinado período, recebeu propina favorecendo 10 dessas empresas com "auditorias de fachada". A firma contratada pelo governo decide realizar novas auditorias em 19 delas, selecionadas aleatoriamente. Seja x o número de empresas, dentre essas 19, onde a firma confirma a existência de irregularidades.

7. Um país A, armado nuclearmente, dispara acidentalmente 10 mísseis intercontinentais contra o país B. Desses 10 mísseis, 7 estão carregados com ogivas nucleares. O país B responde com 7 mísseis de defesa para derrubar pelo menos 7 dos mísseis invasores, porém não conhece qual dos mísseis invasores está ou não carregado com ogivas nucleares. Supondo que os 7 mísseis de defesa acertam precisamente 7 mísseis invasores, seja x o número de mísseis invasores contendo ogivas nucleares que acabam atingindo o país B.

Fim!