Noções de Probabilidade e Estatística - Resolução Exercícios Pares

Gledson Luiz Picharski

August 6, 2007

Capitulo 3

Seção 3.1

Exercício 2

Existem várias respostas para este problema, faço então uma suposição para os itens e calculo as probabilidades.

a)

```
> freq = c(10, 20, 30, 35, 30, 15, 5)
> x <- data.frame(num.filhos = 0:6, freq, prob = freq/sum(freq))
> x

  num.filhos freq       prob
1          0   10 0.06896552
2          1   20 0.13793103
3          2   30 0.20689655
4          3   35 0.24137931
5          4   30 0.20689655
6          5   15 0.10344828
7          6    5 0.03448276
```

b)

```
> freq = c(5, 10, 28, 32, 35, 20, 10, 5)
> x <- data.frame(num.filhos = 0:7, freq, prob = freq/sum(freq))
> x

  num.filhos freq       prob
1          0    5 0.03448276
2          1   10 0.06896552
3          2   28 0.19310345
4          3   32 0.22068966
5          4   35 0.24137931
6          5   20 0.13793103
7          6   10 0.06896552
8          7    5 0.03448276
```

Exercício 4

Devem ser respeitados os axiomas de probabilidade para fazer esta suposição, uma solução possivel seria:

```
> x <- data.frame(anos = 1:5, prob = c(0.3, 0.25, 0.2, 0.15, 0.1))
```

Exercício 6

	anos	prob
1	1	0.30
2	2	0.25
3	3	0.20
4	4	0.15
5	5	0.10

a)

$$f(\boldsymbol{x}) = \begin{cases} 0, sex < 10; \\ 0.2, se 10 \leq x < 12; \\ 0.3, se 12 \leq x < 13; \\ 0.4, se 13 \leq x < 25; \\ 0.1, sex \geq 25. \end{cases}$$

b) $P(X \leq 12) = P(X < 12) + P(X = 12) = 0.2 + 0.3 = 0.5$

c) $P(X < 12) = 0.2$

d) $P(12 \leq X \leq 20) = 0.3 + 7 \times \frac{0.4}{9} = 0.61$

e) $P(X > 18) = 6 \times \frac{0.4}{9} + 0.1 = 0.37$

Seção 3.2

Exercício 2

```
> x <- 1:10
```

a)

```
> mean(x >= 7)
[1] 0.4
```

b)

```
> mean(x <= 7 & x > 3)
[1] 0.4
```

c)

```
> mean(x < 2 | x >= 8)
[1] 0.4
```

d)

```
> mean(x > 5 | x > 8)
[1] 0.5
```

e)

```
> mean(x > 3 & x < 6)
[1] 0.2
```

f)

```
> mean(x[x >= 6] <= 9)
[1] 0.8
```

Exercício 4

Não existe resposta única, mas vou reproduzir possiveis respostas para os problemas.

a) O modelo binomial não parece adequado para está situação, pois existe uma chance diferente de cada aluno ser dorgado, para ser Binomial a chance de cada um ser drogado deferia ser a mesma.

b) Neste caso o modelo Binomial não se aplica, pois cada lâmpada possui chance diferênte de ter defeito.

c) Por serem da mesma fábrica e mesmo modelo, podemos considerar que a chance que cada carro tem de ser poluente é a mesma, então o modelo Binomial poderia ser usado, visto que temos um número limidado de eventos e cada um pode ter apenas como resultado falha ou aprovado.

d) Considerando que o motorista não está ganhando experiência durante o teste, podemos utilizar o modelo Binomial, pois cada evento, possui chance semelhante ao anterior, e temos duas possibilidades para cada evento que são acerto e erro.

Exercício 6

$X \sim Binomial(n, p)$
$n = 15$
$p = 0.8$
$X : \#\ de\ pessoas\ curadas.$

```
> n <- 15
> p <- 0.8
```

a) Basta calcular a Binomial para $X = 15$.

```
> pbinom(14, n, p, lower = F)
[1] 0.03518437
```

b) Pelo menos 13 serão curados, então teremos $X \leq 13$.

```
> pbinom(13, n, p)
[1] 0.8328742
```

c) Isto significa que teremos $X \geq 10$.

```
> pbinom(9, n, p, lower = F)
[1] 0.9389486
```

Seção 3.3

Exercício 2

$X \sim Geomtrica(0.5)$.

a)

```
> pgeom(2, 0.5)
[1] 0.875
```

b)

```
> pgeom(1, 0.5, lower = F)
[1] 0.25
```

c)

```
> sum(dgeom(4:5, 0.5))
[1] 0.046875
```

d)

```
> qgeom(0.8, 0.5)
[1] 2
```

Exercício 4

$X \sim poisson(1)$

a) $P(X \geq 1)$

```
> ppois(0, 1, lower = F)
```

```
[1] 0.6321206
```

b) $P(X \leq 2)$

```
> ppois(2, 1)
[1] 0.9196986
```

c) $P(2 \leq X \leq 4)$

```
> sum(dpois(2:4, 1))
[1] 0.2605813
```

d) $P(X \leq 1)$

```
> ppois(1, 1)
[1] 0.7357589
```

Exercício 6

Observamos que neste caso para o R, n vai ser o número de peças boas e m o número de defeituosas.

a)

```
> phyper(2, 3, 9, 4, lower = F)
[1] 0.01818182
```

b)

```
> phyper(4, 3, 9, 4)
[1] 1
```

c)

```
> phyper(3, 3, 9, 4)
[1] 1
```

Seção 3.4

Exercício 2

Faço algumas simulações de valores para conseguir os dados que o livro deu de forma resumida, então obtenho a função de probabilidade e a função densidade.

```
> transporte <- matrix(c(3,3.5,4,0.5,0.3,0.2),ncol=2)
> estadia <- matrix(c(2,2.5,3,3.5,rep(0.25,4)),ncol=2)
> despesa <- matrix(c(rep(transporte[,1],each=4)+rep(estadia[,1],3),
+                   rep(transporte[,2],each=4)*rep(estadia[,2],3)),ncol=2,
+                   dim =  list(rep("",12),c("valor","prob")))
> ##função de probabilidade da despesa de viagem.
> despesa
 valor  prob
   5.0 0.125
   5.5 0.125
   6.0 0.125
   6.5 0.125
   5.5 0.075
   6.0 0.075
   6.5 0.075
   7.0 0.075
   6.0 0.050
   6.5 0.050
   7.0 0.050
   7.5 0.050
> for(i in 1:12){a <- paste("a",1:12,sep=""); assign(a[i],sum(despesa[1:i,2]))}
> distribuição.despesa <- matrix(c(despesa[,1],unlist(lapply(a,get))),
+                                ncol=2,dim =  list(rep("",12),c("valor","prob")))
> ##função distribuição de probabilidade da despesa de viagem.
> distribuição.despesa
```

```
valor  prob
  5.0 0.125
  5.5 0.250
  6.0 0.375
  6.5 0.500
  5.5 0.575
  6.0 0.650
  6.5 0.725
  7.0 0.800
  6.0 0.850
  6.5 0.900
  7.0 0.950
  7.5 1.000
```

Exercício 4

O número 1 representa a ocorrência de falha na parte descrita, e 0 representa a não ocorrência de falha, as três primeiras colunas de prob são correspondente ao aparecimento ou não de falha, a coluna prob é obtida pelo produto das três probabilidades, visto que são eventos independentes.

```
> x <- data.frame(ele = rep(c(0, 1), each = 4), mec = rep(c(0,
+     1, 0, 1), each = 2), est = rep(c(0, 1), 4), prob1 = rep(c(0.9,
+     0.1), each = 4), prob2 = rep(c(0.9, 0.1, 0.9, 0.1), each = 2),
+     prob3 = rep(c(0.9, 0.1), 4))
> x$prob <- with(x, prob1 * prob2 * prob3)
> x$tempo <- with(x, ele * 10 + mec * 20 + est * 50)
> x$tempo[x$ele == 1 & x$mec == 1] <- x$tempo[x$ele == 1 & x$mec ==
+     1] + 20
> x
```

	ele	mec	est	prob1	prob2	prob3	prob	tempo
1	0	0	0	0.9	0.9	0.9	0.729	0
2	0	0	1	0.9	0.9	0.1	0.081	50
3	0	1	0	0.9	0.1	0.9	0.081	20
4	0	1	1	0.9	0.1	0.1	0.009	70
5	1	0	0	0.1	0.9	0.9	0.081	10
6	1	0	1	0.1	0.9	0.1	0.009	60
7	1	1	0	0.1	0.1	0.9	0.009	50
8	1	1	1	0.1	0.1	0.1	0.001	100

a) Basta somarmos as probabilidades onde onde tivermos o tempo menor que 25 minutos.

```
> with(x, sum(prob[tempo < 25]))
[1] 0.891
```

b) Soma-se as probabilidades correspondentes a tempos maiores que 40 minutos.

```
> with(x, sum(prob[tempo > 40]))
[1] 0.109
```

Exercício 6

a) A função de probabilidade pode ser obtida observando a probabilidade de cada classe mostrada pelo exercício.

$$f(\boldsymbol{x}) = \begin{cases} 0, se x < -1; \\ 0.2, se -1 \leq x < 2; \\ 0.3, se 2 \leq x < 5; \\ 0.2, se 5 \leq x < 6; \\ 0.2, se 6 \leq x < 15; \\ 0.1, se x \geq 15. \end{cases}$$

b) $P(X \leq -2) = 0$

c) $P(X < 2) = 0.2$

d) $P(3 \leq X \leq 12) = 2 \times \frac{0.3}{3} + 0.2 + 6 \times \frac{0.2}{9} = 0.533$

e) $P(X > 14) = 1 * \frac{0.2}{9} + 0.1 = 0.122$

Exercício 8

Aqui podemos observar um caso particular de Bernoulli, pois em cada tentativa só pode dar certo ou errado, mas como temos 4 eventos possiveis, em que o experimento termina quando o macaco acerta na tarefa, e cada tentativa tem probabilidade diferente da anterior, então este modelo é desconhecido, mas podemos dizer que as chances de acertividade são cada veis maiores conforme almentam o número de tentativas.

Exercício 10

Usando o número 0 para representar o sexo feminino e 1 para o masculino, construo uma matriz de casos possiveis e através disto obtenhos as probabilidades.

```
> x <- data.frame(A = rep(c(0, 1), each = 4), B = rep(c(0, 1, 0,
+     1), each = 2), C = rep(c(0, 1), 4), prob = rep(0.5^3, 8))
```

a) Percebemos esta situação onde a soma entre as colunas A, B e C resultam em exatamente 2.

```
> with(x, sum(prob[A + B + C == 2]))
[1] 0.375
```

b) Isso pode ser observado nas linhas onde a soma das três primeiras colunas é maior ou igual a 1.

```
> with(x, sum(prob[A + B + C >= 1]))
[1] 0.875
```

c) Esta situação é observada onde a soma das colunas resulta em exatamente 0.

```
> with(x, sum(prob[A + B + C == 0]))
[1] 0.125
```

Exercício 12

temos 500 unidades deste equipamento, serão retirados 5 para inspeção, sabemos que 10 são defeituosos e 490 são bons, então basta colocar os dados em uma hypergeométrica que obtemos o resultado, assim verifico a probabilidade de haver um equipamento defeituoso entre os 5 retirados.

```
> phyper(1, 490, 10, 5)
[1] 4.041299e-07
```

Exercício 14

O calculo pode ser feito usando a fómula, mas de forma mais simples uso uma função do R para obter a resposta.

a)

```
> ppois(0, 5)
[1] 0.006737947
```

b)

```
> ppois(0, 5, lower = F)
[1] 0.993262
```

c) queremos verificar $P(2 < X \leq 5)$, usa-se a mesma função.

```
> diff(c(ppois(2, 5), ppois(5, 5)))
[1] 0.4913086
```

Exercício 16

$X \sim Binom(n, p)$
$n = 20$
$p = 0.7$
$X : \#\ de\ pacientes\ imunizados$

a) Podemos usar o lower.tail do pbinom para para calcular a probabilidade de ser acima do valor, ou então podemos usar o complementar da probabilidade de ser abaixo do valor. Talvez existam outras maneiras interessantes de resolver esta questão.

```
> pbinom(17, 20, 0.7, lower = F)
[1] 0.03548313
> 1 - pbinom(17, 20, 0.7)
[1] 0.03548313
```

b)

```
> pbinom(4, 20, 0.7)
[1] 5.550253e-06
```

c)

```
> pbinom(17, 20, 0.7)
[1] 0.9645169
```

Exercício 18

$X \sim Binom(n, p)$
$n = 12$
$p = 0.8$
$X : \#\ de\ pacintes\ completamente\ curados.$

a) $P(X = 8)$

```
> dbinom(8, 12, 0.8)
[1] 0.1328756
```

b) $P(X < 3, X > 5)$

```
> sum(c(pbinom(2, 12, 0.8), pbinom(5, 12, 0.8)))
[1] 0.003907658
```

c) $P(X < 10)$

```
> pbinom(9, 12, 0.8)
[1] 0.4416543
```

Exercício 20

$X \sim Pois(\lambda)$
$\lambda = 1$
$X : \#\ de\ avies\ por\ minuto.$

a) $P(X = 3)$

```
> dpois(3, 1)
[1] 0.06131324
```

b) Para que algum avião fique sem atendimento imediato já deve ter dois sendo atendidos, ou seja, temos que calcular $P(X > 2)$

```
> ppois(2, 1, lower = F)
[1] 0.0803014
```

c) Com o tráfego dobrando, teremos λ sendo 2, sendo 3 aviões a nova capacidade, devemos calcular: $P(X > 3)$

```
> ppois(3, 2, lower = F)
[1] 0.1428765
```

Exercício 22

$X \sim Pois(\lambda)$
$\lambda = 4$
$X : \#\ de\ acessos\ \ CPU\ por\ segundo.$

a) Queremos calcular: $P(X > 2)$ e $P(X \leq 5)$

```
> ppois(2, 4, lower = F)
[1] 0.7618967
```

```
> ppois(5, 4)
[1] 0.7851304
```

b) Para $\lambda = 10$, queremos: $P(X = 50)$

```
> dpois(50, 10)
[1] 1.492727e-19
```

Exercício 24

$X \sim Geomtrica(p)$
$p = 0.8$

$$f(y) = \begin{cases} 0, sey < 0; \\ 0.8\dot{(}0.2)^y, se0 \leq y < 6; \\ 0.8\dot{(}0.2)^6, sey \geq 6. \end{cases}$$

```
> df <- function(y) {
+     if (y < 0)
+         0
+     else {
+         if (y >= 0 | y < 6)
+             dgeom(y, 0.8)
+         else {
+             dgeom(6, 0.8)
+         }
+     }
+ }
> pf <- function(y) {
+     sum(sapply(0:y, df))
+ }
```

a) Poderiamos observar a função de probabilidade para ver a condição e então calcular usando a função dpois(q,p), ou podemos usar a função que criei.

```
> df(2)
[1] 0.032
```

b) A probabilidade acumulada até o ponto 2.5 pode ser obtida por: $P(X < 3)$

```
> pf(3)
[1] 0.9984
```

c) $\dfrac{P(Y = 3 \cap Y \leq 5)}{P(Y <= 5)} = \dfrac{P(X = 3}{P(Y <= 5)}$

```
> df(3)/pf(5)
[1] 0.00640041
```

d) $P(3 \leq Y \leq 5)$

```
> diff(c(pf(2), pf(5)))
[1] 0.007936
```

Exercício 26

Temos aqui, uma hipergeométrica, com m=4, n= 5 e r = 3, sendo o total de jacarés m+n.

a)

```
> dhyper(3, 4, 5, 3)
[1] 0.04761905
```

b) $P(X \geq 1)$

```
> phyper(0, 4, 5, 3, lower = F)
[1] 0.8809524
```

c) $P(X \geq 2)$

```
> phyper(1, 4, 5, 3, lower = F)
```

```
[1] 0.4047619
```

Exercício 28

A tabela pode ser obtida com o diretamente do site, ou de um arquivo local, basta direcionar corretamente.

```
> tab <- read.table("http://www.ime.usp.br/~noproest/dados/questionario.txt",
+     head = T)
> head(tab)
```

```
  Id Turma Sexo Idade  Alt Peso Filhos Fuma Toler Exerc Cine OpCine TV OpTV
1  1     A    F    17 1.60 60.5      2  NAO     P     0    1      B 16    R
2  2     A    F    18 1.69 55.0      1  NAO     M     0    1      B  7    R
3  3     A    M    18 1.85 72.8      2  NAO     P     5    2      M 15    R
4  4     A    M    25 1.85 80.9      2  NAO     P     5    2      B 20    R
5  5     A    F    19 1.58 55.0      1  NAO     M     2    2      B  5    R
6  6     A    M    19 1.76 60.0      3  NAO     M     2    1      B  2    R
```

a) Verifico o minimo e o máximo, então escolho o tamanho de cada classe e divido em classe.

```
> with(tab, range(Exerc))
```

```
[1]  0 10
```

```
> with(tab, table(cut(Exerc, c(0:5 * 2), right = F, include = T)))
```

```
 [0,2)  [2,4)  [4,6)  [6,8) [8,10] 
    11     14     12      8      5 
```

b) A situação corresponde ao modelo hipergeométrico, onde podemos dividir casos favoraveis como sendo 6 horas ou mais de exercicios, e desfavoraveis o complementar disto.

```
> m <- with(tab, sum(Exerc >= 6))
> n <- with(tab, sum(Exerc < 6))
> dhyper(3, m, n, 5)
```

```
[1] 0.08989975
```

c)

```
> dhyper(5, m, n, 5)
```

```
[1] 0.0006074308
```

Exercício 30

```
> se <- read.xls("aeusp.xls", head = T)
> head(se)
```

```
  Num    Comun Sexo Idade Ecivil X.Reproce X.Temposp X.Resid Trab Ttrab X.Itrab
1   1 JdRaposo    2     4      4  Nordeste        21       9    3    NA      20
2   2 JdRaposo    2     1      1   Sudeste        24       9    1     1      14
3   3 JdRaposo    2     2      1  Nordeste        31       3    1     1      14
4   4 JdRaposo    1     2      2  Nordeste        10       3    1     4      10
5   5 JdRaposo    2     4      2  Nordeste        31       6    1     1      11
6   6 JdRaposo    2     4      2   Sudeste        24       4    2    NA      15
  X.Renda X.Acompu X.Serief
1       1        2        1
2       2        2        7
3       5        2        7
4       5        2       11
5       6        1        4
6       4        2        4
```

```
> names(se) <- lapply(strsplit(names(se), "X."), function(x) x[length(x)])
> head(se)
```

```
  Num    Comun Sexo Idade Ecivil  Reproce Temposp Resid Trab Ttrab Itrab Renda
1   1 JdRaposo    2     4      4 Nordeste      21     9    3    NA    20     1
2   2 JdRaposo    2     1      1  Sudeste      24     9    1     1    14     2
3   3 JdRaposo    2     2      1 Nordeste      31     3    1     1    14     5
4   4 JdRaposo    1     2      2 Nordeste      10     3    1     4    10     5
5   5 JdRaposo    2     4      2 Nordeste      31     6    1     1    11     6
6   6 JdRaposo    2     4      2  Sudeste      24     4    2    NA    15     4
  Acompu Serief
1      2      1
2      2      7
3      2      7
4      2     11
5      1      4
6      2      4
```

```
> with(se, Sexo[Sexo != 1 & Sexo != 2] <- NA)
> with(se, Idade[Idade < 1 | Idade > 4] <- NA)
> with(se, Ecivil[Ecivil < 1 | Ecivil > 5] <- NA)
> with(se, Temposp[Temposp[Idade == 1] > 25] <- NA)
> with(se, Temposp[Temposp[Idade == 2] > 35] <- NA)
> with(se, Temposp[Temposp[Idade == 3] > 45] <- NA)
> with(se, Temposp[Temposp[Idade == 4] > Inf] <- NA)
> with(se, Idade[Temposp == NA] <- NA)
> with(se, Trab[Trab < 1 | Trab > 3] <- NA)
> with(se, Ttrab[Ttrab < 1 | Ttrab > 5] <- NA)
> with(se, Renda[Renda < 1 | Renda > 6] <- NA)
> with(se, Acompu[Acompu < 1 | Acompu > 2] <- NA)
> with(se, Serief[Serief < 1 | Serief > 12] <- NA)
```

a) Observamos na Figura 1 que os dados parecem com a distribuição de poisson, e faz sentido, pois mostra uma relação com a faixa de tempo que as pessoas moram na região.

```
> with(se, range(Temposp))
```

```
[1]  1 67
```

```
> with(se, table(cut(Temposp, 0:7 * 10)))
```

```
 (0,10] (10,20] (20,30] (30,40] (40,50] (50,60] (60,70]
    105     124     105      29      19       2       1
```

```
> with(se, hist(Temposp, breaks = 0:7 * 10, main = ""))
```

b)

```
> R.obs <- with(se, table(Resid))
> R.obs <- c("0" = 0, R.obs)
> R.esp <- dbinom(0:10, size = 10, p = 0.5)
```

```
> barplot(rbind(R.obs/sum(R.obs), R.esp), bes = T, col = c("gray70",
+     "gray30"))
> legend("topright", c("Observados", "Esperados"), fill = c("gray70",
+     "gray30"))
```

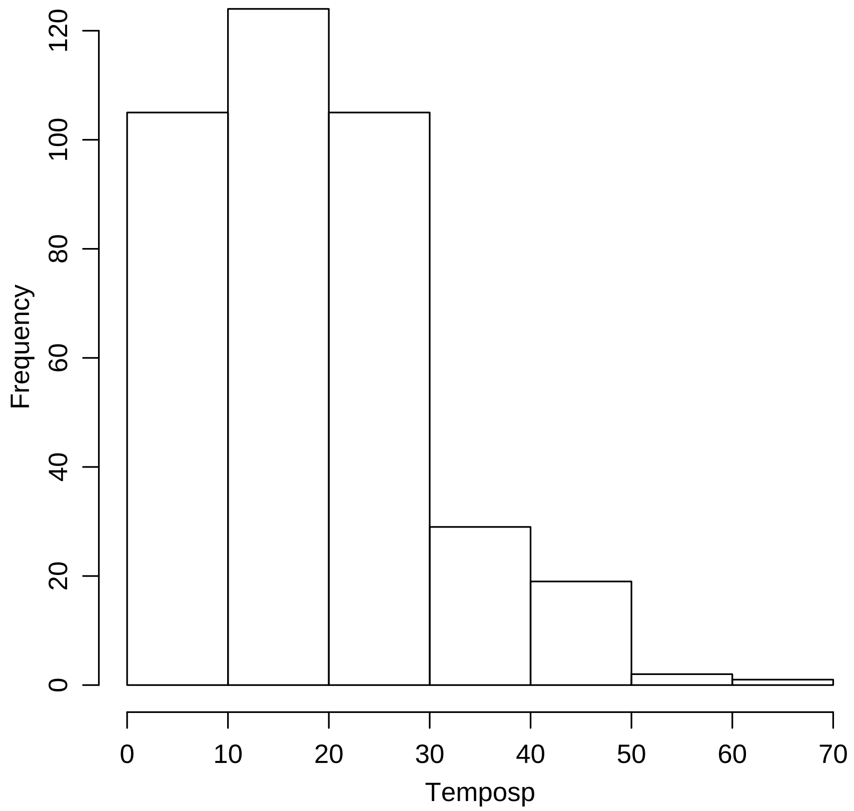

Figure 1: Histograma do tempo de residencia em SP

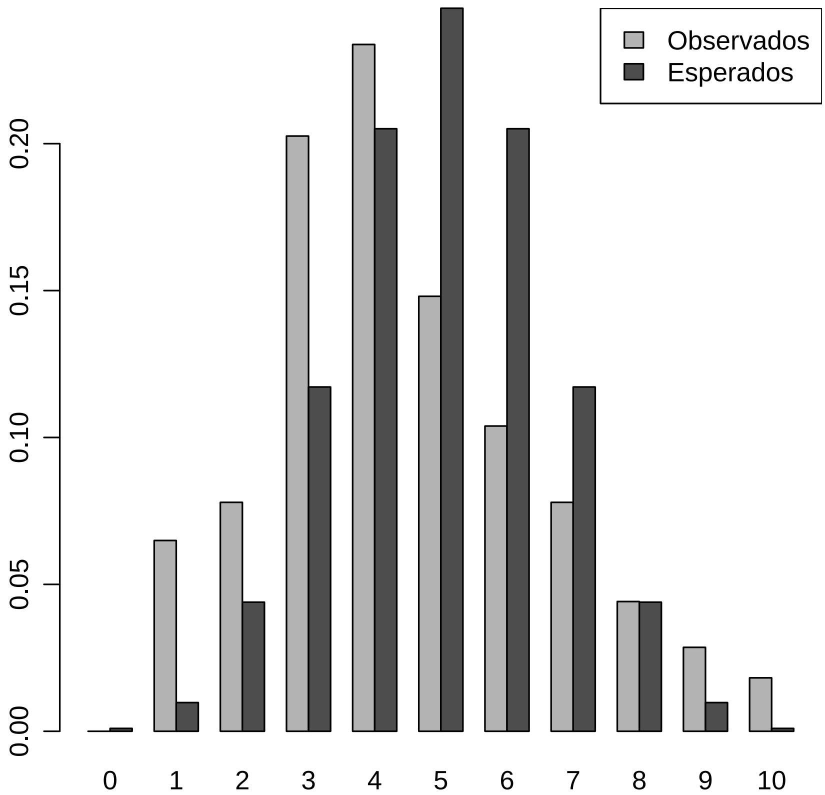

Figure 2: Comparativo entre o Modelo Binomial e a variavel número de residentes(Resid).