

1ª Lista de Exercícios

Eniedson Fabiano Pereira da Silva Junior - 119110512
João Paulo Alves Dantas - 119210258 Rodrigo Farias Oliveira - 118110711

Respostas

1. Ao ter uma grande quantidade de dados com valores distintos para se exibir em uma tabela pode ser que o resultado gerado a não seja tão atrativo, como visto na tabela a seguir, que detalha a quantidade de horas extras dormidas por 20 pacientes que consumiram duas categorias de drogas soníferas:

```
##
##      1 2
## -1.6 1 0
## -1.2 1 0
## -0.2 1 0
## -0.1 1 1
##  0   1 0
##  0.1 0 1
##  0.7 1 0
##  0.8 1 1
##  1.1 0 1
##  1.6 0 1
##  1.9 0 1
##  2   1 0
##  3.4 1 1
##  3.7 1 0
##  4.4 0 1
##  4.6 0 1
##  5.5 0 1
```

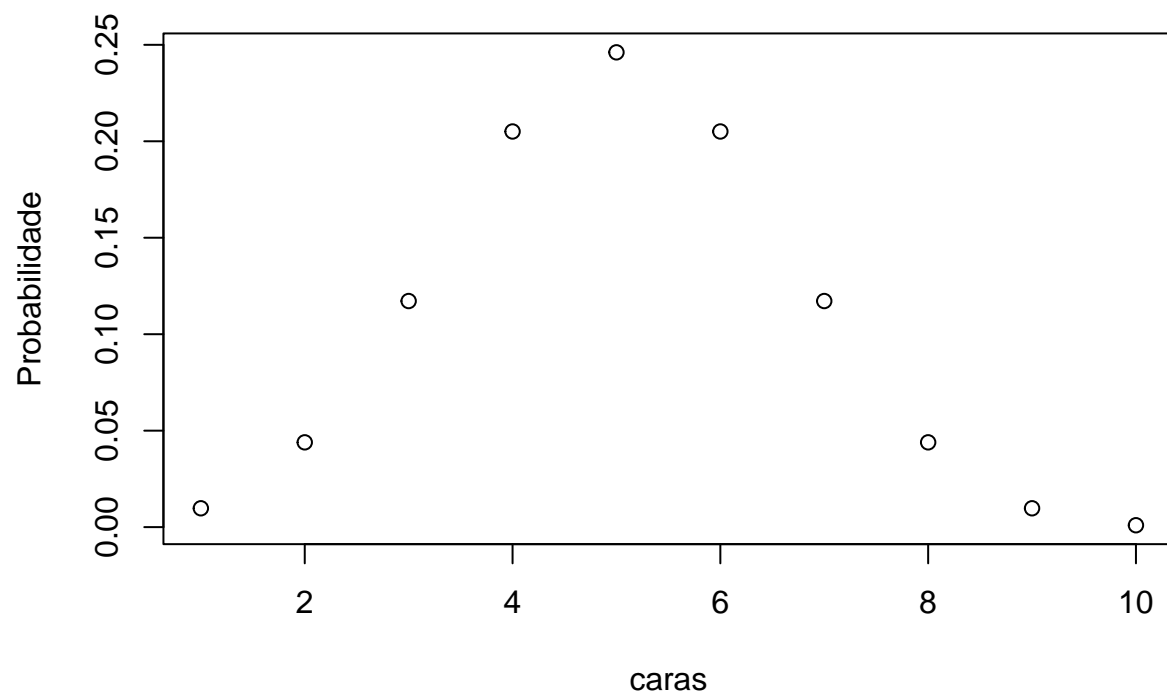
Porem podemos agrupar os resultados em intervalos de valores arbitrários, tornando uma visualização bem mais atrativa:

```
##
## sono      1 2
## -2..0 5 1
##  0..2 3 5
##  2..5 2 3
```

2. Uma variável aleatória pode ser definida intuitivamente como um evento que possui chances (iguais ou não) de assumir determinados valores, como, por exemplo, o lançamento de um dado balanceado, aonde cada face do dado possui chance igual a $1/6$ de cair, ou até mesmo o lançamento de uma moeda justa, que possui chance de $1/2$ para cada face.

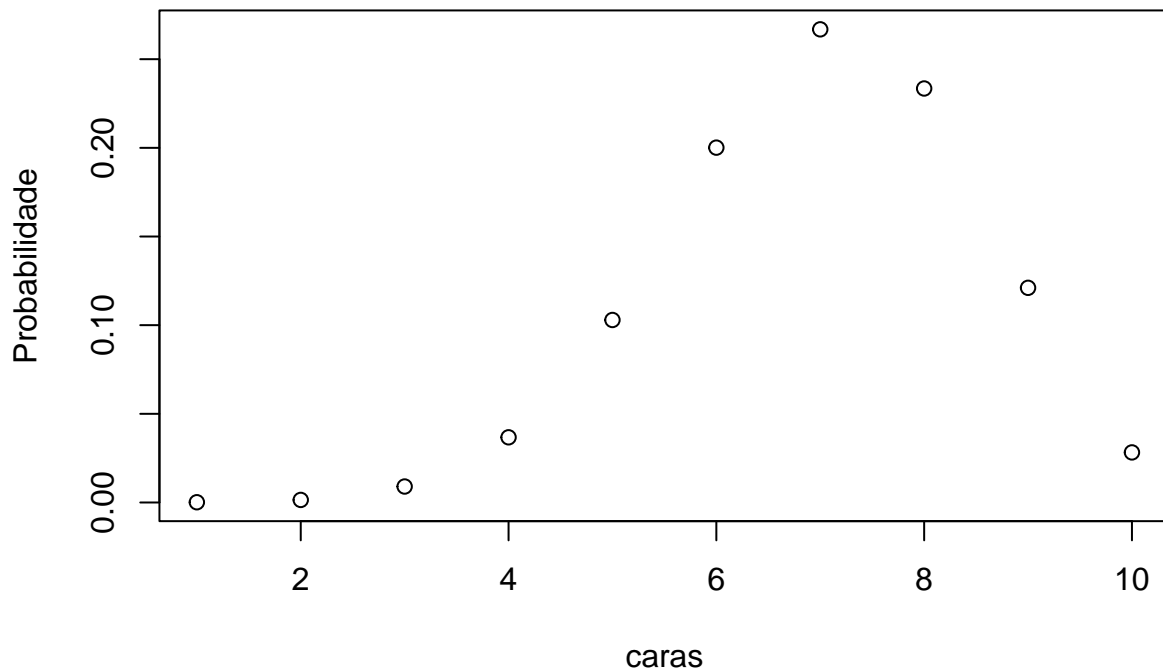
Para visualizar a manipulação das probabilidades temos aqui um gráfico que mostra a probabilidade de conseguir x caras ao lançar uma moeda justa:

Chance de conseguir x caras ao lançar 10 moedas balanceada



Agora ajustando a moeda para ela ter uma chance de 7/10 de cair cara, vemos que o gráfico se move uma probabilidade maior de conseguir altos números de caras:

Chance de conseguir x caras ao lançar 10 moedas não justa



3. As variáveis aleatórias podem ser divididas em dois grupos principais, qualitativas e quantitativas, dentro desses podemos ainda dividir cada um em outros dois, temos as qualitativas nominais e ordinais, e temos as quantitativas discretas e contínuas.

3.1. Qualitativa Nominal: Categoria de variável em que é impossível mensurar ou ordenar os valores, como no exemplo abaixo em que temos os sobreviventes do Titanic agrupados por seu gênero, que por sua vez é uma variável aleatória.

```
##           Sex
## Survived Male Female
##      No  1364    126
##      Yes   367    344
```

3.2. Qualitativa Ordinais: Nessa categoria de variável é possível atribuir uma ordem aos valores coletados. Utilizando a mesma base de dados, podemos agora observar os sobreviventes do Titanic agrupados pela classe em que embarcaram, podendo ser 1.^a Classe, 2.^a Classe, 3.^a Classe ou tripulação.

```
##           Class
## Survived 1st 2nd 3rd Crew
##      No  122 167 528  673
##      Yes  203 118 178  212
```

3.3. Quantitativas Discretas: São variáveis aleatórias quantitativas discretas aquelas que podem ser contadas e que possuem um número finito de possíveis valores. Como exemplo temos o número de crianças em uma casa, quantos carros uma família possui, número de internações em um hospital, etc.

3.4. Quantitativas Contínuas: São dados que de certa forma possuem infinitas possibilidades de valores. Por exemplo, ao medir o peso de uma pessoa, dependendo da acurácia da balança, é possível chegar a um resultado quase que único para cada indivíduo, a mesma coisa para a idade de uma pessoa ou sua

altura, por esse motivo geralmente as variáveis Quantitativas Contínuas são geralmente arredondadas para diminuir a quantidade de possíveis valores.

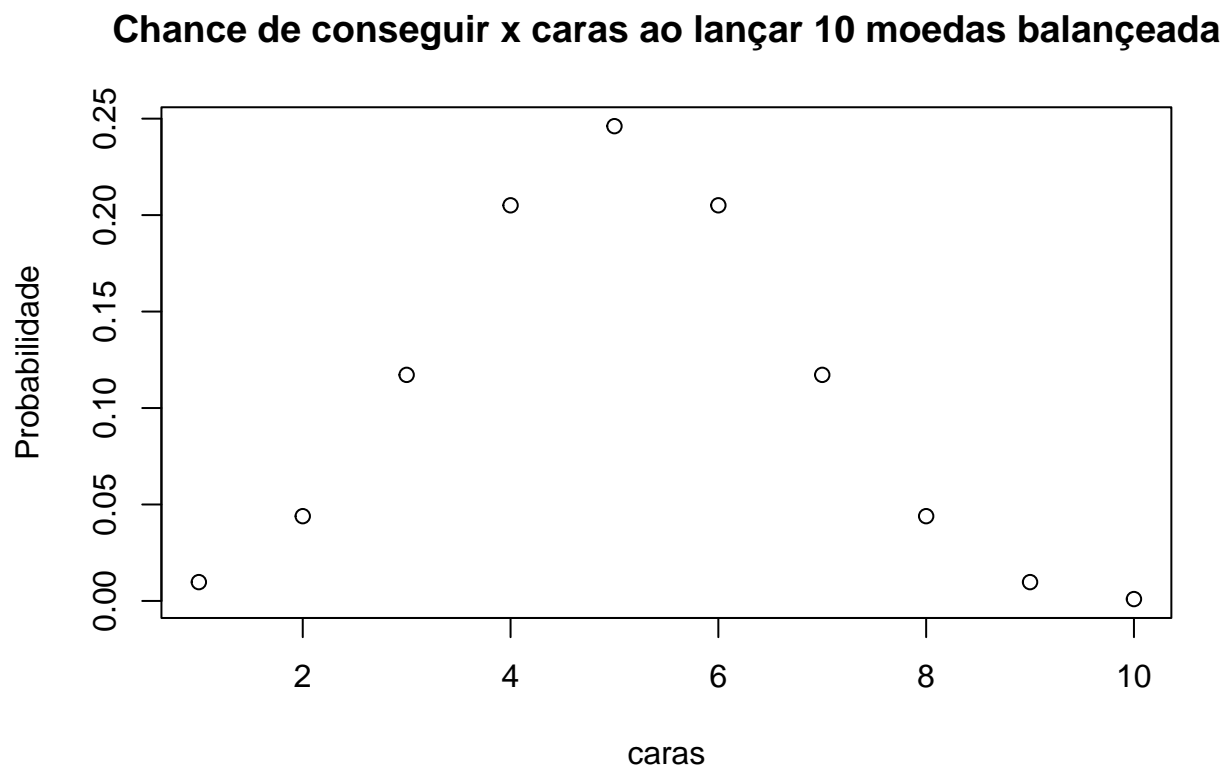
4. Uma função probabilidade mapeia os possíveis resultados de uma variável aleatória com base nos seus parâmetros aleatórios. Como exemplo podemos fazer utilizar o mesmo da questão 2, agora escrevendo a função probabilidade de retirar k caras em um total de n lançamentos com uma probabilidade p de sucesso:

$$P = \binom{n}{k} p^k (1-p)^{n-k}$$

Com uma moeda balanceada e 10 lançamentos chegamos a seguinte função:

$$P = \binom{10}{k} 0.5^k (0.5)^{10-k}$$

Que resulta no mesmo gráfico visto na questão 2:



5. Uma Distribuição de Probabilidades é um par de probabilidade e valor que pode ocorrer em um evento, como, por exemplo, um dado que possui chances diferentes de cair em cada lado. Como, por exemplo, na tabela a seguir.

```
##          1    2    3    4    5    6
## chance 0.1 0.2 0.3 0.2 0.1 0.1
```

6. De certa maneira sim, às duas distribuições associam valores a eventos que ocorrem, por um lado a distribuição de probabilidade atribui chances de que um determinado evento ocorra dentro de um espaço amostral, já a distribuição de frequências é algo mais empírico e tenta chegar nesse mesmo resultado

recordando registros. Em espaços amostrais grandes às duas tendem a chegar em um resultado muito parecido, quando analisando um mesmo evento. Um bom exemplo dessa semelhança pode ser vista analisando o cara-ou-coroa, em que uma Distribuição de Probabilidades define que uma moeda justa tem 1/2 de chance de cair em cada lado, porém em um teste empírico o matemático inglês John Kerrich jogou uma moeda 10 mil vezes, registrando 5.067 caras e 4.933 coroas, números muito próximos dos 50%, porém com uma pequena discrepância.

7. Uma **Distribuição Bernoulli** é uma distribuição discreta que possui espaço amostral $\{0, 1\}$, onde 0 representa a probabilidade de falha e o valor 1 representa a probabilidade de sucesso. Por exemplo:
 - 7.1. Estudar as probabilidades referentes aos atrasos em voos comerciais, onde o sucesso seria chegar na hora ou adiantado e a falha seria perder o voo.
 - 7.2. Estudar as probabilidades de aprovação e reprovação em uma determinada disciplina do curso, onde o sucesso seria a aprovação na disciplina e a falha a reprovação na mesma.
8. É necessário se conhecer o parâmetro \mathbf{p} , que representa a probabilidade de sucesso perante o caso estudado.
9. Uma distribuição Bernoulli segue a seguinte f.p. onde k representa a se a probabilidade é de sucesso ou falha, sendo 1 para sucesso e 0 para falha, e p é a probabilidade de sucesso:

$$P(X = k) = p^k(1 - p)^{1-k}$$

9.1. Um exemplo para uso dessa fórmula pode ser encontrado no seguinte enunciado, junto da sua solução: Qual a probabilidade de se acertar uma questão “no chute” sabendo que a questão possui 4 alternativas de escolha e apenas uma delas é correta?

Para esse caso, temos que $k = 1$ e $p = 0,25$, substituindo na fórmula, temos:

$$P(X = 1) = 0,25^1 * (1 - 0,25)^{1-1}$$

$$P(X = 1) = 0,25 * (1 - 0,25)^0$$

$$P(X = 1) = 0,25 * 1 = 0,25 = 25\%$$

9.2. Usando o mesmo exemplo, para descobrir a probabilidade da falha ocorrer, basta substituir o k por 0, dessa forma, temos:

$$P(X = 0) = 0,25^0 * (1 - 0,25)^{1-0}$$

$$P(X = 0) = 1 * (0,75)^1 = 0,75 = 75\%$$

10. Uma Distribuição Binomial pode ser usada quando precisamos descobrir as probabilidades a partir da repetição de determinado experimento e estudar os possíveis resultados referentes ao número de sucessos dentre os experimentos. Um exemplo simples da utilização de uma técnica como essa se dá quando precisamos, por exemplo, saber a probabilidade de, entre 5 produtos, 3 serem defeituosos, dessa forma, seria uma distribuição binomial onde $n = 5$ e $k = 3$.
11. Deve se conhecer o parâmetro \mathbf{p} , que se refere a probabilidade de dar certo e o parâmetro \mathbf{n} , que se refere ao número total de experimentos que serão realizados.
12. Uma distribuição Binomial segue a seguinte f.p. onde k representa a quantidade de sucessos, n representa o número de repetições do experimento e p é a probabilidade de sucesso:

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

12.1. Um exemplo para uso dessa fórmula pode ser encontrado no seguinte enunciado, junto da sua solução: Qual a probabilidade de se acertar 3 questões “no chute” entre 4 questões, no total, sabendo que cada questão possui 4 alternativas de escolha e apenas uma delas é correta? Para esse caso, temos que $k = 3$, $n = 4$ e $p = 0,25$, substituindo na fórmula, temos:

$$P(X = 3) = \binom{4}{3} * 0,25^3 * (1 - 0,25)^{4-3}$$

$$P(X = 3) = \frac{4!}{3!1!} * 0,015625 * 0,75$$

$$P(X = 3) = 0,046875 = 4,6875\%$$

13. Em resumo, a função densidade de probabilidade se trata de uma equação capaz de representar a distribuição das chances de uma variável contínua assumir determinados valores. Como por exemplo, pode-se usá-la para observar as faixas etárias dos funcionários de uma empresa e as chances de alguém necessitar de maiores cuidados, como acontece quando se têm pessoas mais velhas engajadas dentro da mesma.
14. A quantidade de produtos que saem com defeito por hora em uma fábrica, a faixa etária de internações por covid em determinada região, a quantidade de acidentes sofridos por dia da semana em alguma cidade, entre vários outros fenômenos aleatórios.
15. Os parâmetros primordiais para o cálculo da Distribuição Normal são a Média (representada pela letra grega μ), que indica o valor médio das ocorrências, e o desvio padrão da distribuição (representado pela letra σ , e que pode ser obtido através da raiz quadrada da variância).
16. “Em uma empresa a renda média mensal é R\$ 8000,00 e o desvio-padrão é R\$ 2000,00. Considerando que a renda mensal seja normalmente distribuída. Determinar a probabilidade de um trabalhador ganhar entre R\$ 7500,00 e R\$ 9000,00 por mês.”

Utilizando-se da f.d.p. da Distribuição normal, dada por “ $Z = (x - \mu) / \sigma$ ”, temos o seguinte: $Z1 = (7500 - 8000)/2000 = -500/2000 = -0,25$ $Z2 = (9000 - 8000)/2000 = 1000/2000 = 0,50$

Logo, pela Tabela Normal Padrão, temos a probabilidade de 9,871%, para Z1, e de 19,146% para Z2. Que totaliza uma probabilidade final de 29,017% de um trabalhador receber um salário entre R\$ 7500,00 e R\$ 9000,00.

17. Podemos dividir uma pesquisa estatística nas seguintes fases:
 - Definição do problema: onde o pesquisador deve formular de forma concisa o problema alvo;
 - Planificação do processo de resolução: na qual o pesquisador determina o processo que será utilizado para resolução e coleta dos dados em questão, decidindo também entre a observância de uma população ou de uma amostra;
 - Recolha de dados: fase onde se coleta os dados desejados através de algum meio escolhido pelo pesquisador (questionários, entrevistas, observação, etc);
 - Organização de dados: fase que converte e organiza os dados coletados, em verdadeiras informações agrupadas;
 - Apresentação de dados: onde ocorre a sintetização das informações em tabelas ou gráficos, para que se consiga apresentar o grande volume coletado, de forma facilitada;
 - Análise e interpretação de dados: fase final em que se calculam novas informações com base nos dados estatísticos, afim de obter uma conclusão/explicação para o fenômeno observado na pesquisa em questão.