

Oficina de Estatística Tutorial 3

Virgilio Mendes

16/07/2019

Tutorial 3

Neste tutorial, vamos trabalhar a ideia de probabilidade, para fundamentar a sua importância na produção de inferências estatísticas; amostragem, para destacar elementos de uma distribuição normal.

Uma maneira simples e comum de se demonstrar como funcionam algumas ideias de probabilidade, amostragem e distribuição normal é a simulação de resultados de um dado. Sendo assim, vamos simular um dado de 6 faces no R para mostrar como isso pode funcionar.

Primeiramente, criamos um vetor chamado “dado” que terá as faces 1 a 6:

```
dado <- 1:6
```

1. Probabilidade

Qual a probabilidade de que uma das 6 faces seja sorteada ao lançarmos o dado? Ela é de 1/6.

Sendo assim, dada a **independência** de um lançamento de dados (ou seja, o resultado de uma jogada não depende da jogada anterior), esperamos que em um número alto de repetições ou de observações, as ocorrências de cada resultado do dado se aproxime da sua probabilidade. Vamos testar?

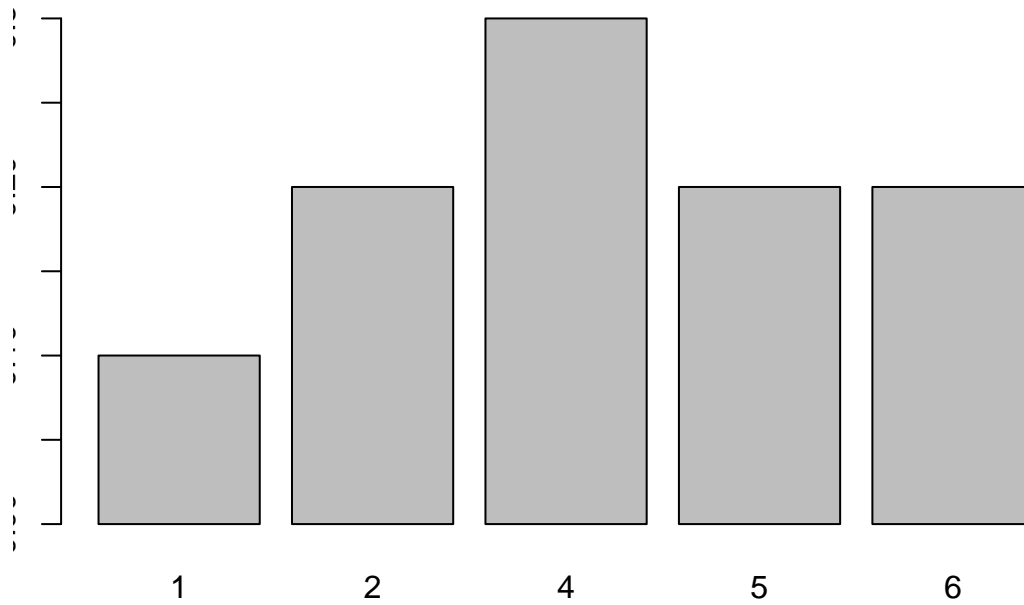
Inicialmente, vamos simular 10 lançamentos de dados e apresentar este resultado graficamente.

```
set.seed(1234)

lancamento_dados <- sample(dado, size = 10, replace = T)

lancamento_dados <- prop.table(table(lancamento_dados))

barplot(lancamento_dados)
```



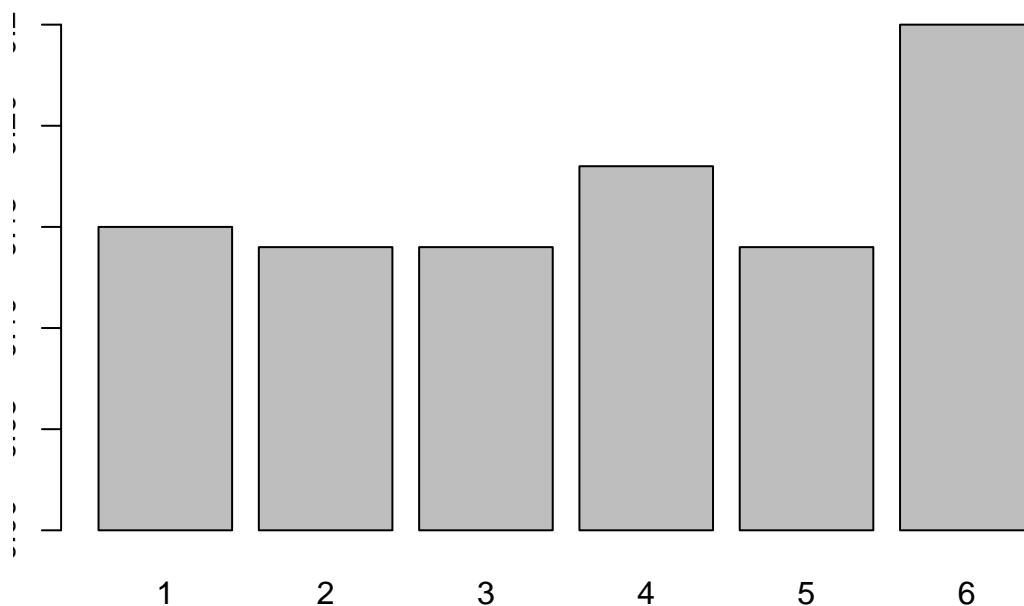
Com esses 10 lançamentos, vemos que o número 3 ainda não foi sorteado, e o número de vezes que cada número foi sorteado ainda não está balanceado. E se aumentarmos o número de lançamentos para 100?

```
set.seed(1234)

lancamento_dados <- sample(dado, size = 100, replace = T)

lancamento_dados <- prop.table(table(lancamento_dados))

barplot(lancamento_dados)
```



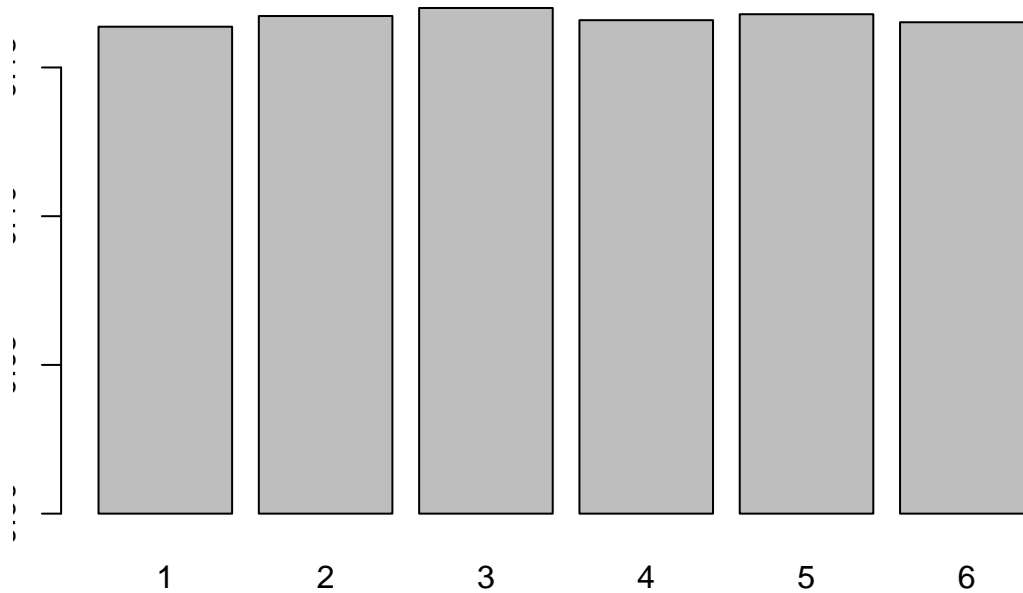
Os resultados do dado vão se aproximando de uma uniformidade, mais ainda não a alcançaram. E se aumentarmos o número de lançamentos para 10.000?

```
set.seed(1234)

lancamento_dados <- sample(dado, size = 10000, replace = T)

lancamento_dados <- prop.table(table(lancamento_dados))

barplot(lancamento_dados)
```



Com 10.000 lançamentos, o número de ocorrências ainda não é idêntico ao da probabilidade, mas se encontra bastante próxima. Vejamos a tabela de frequência relativa desses lançamentos:

```
lancamento_dados
```

```
lancamento_dados 1 2 3 4 5 6 0.1637 0.1673 0.1700 0.1659 0.1679 0.1652
```

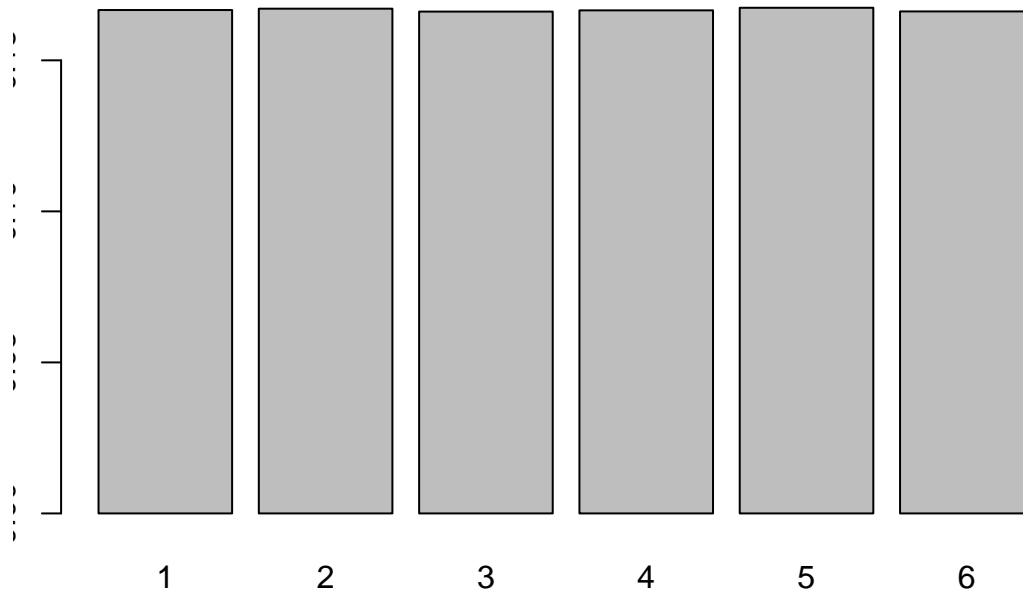
E com 1.000.000 de lançamentos de dados?

```
set.seed(1234)

lancamento_dados <- sample(dado, size = 1000000, replace = T)

lancamento_dados <- prop.table(table(lancamento_dados))

barplot(lancamento_dados)
```



Agora ficamos ainda mais próximos da probabilidade de ocorrência de uma das 6 faces do dado.

2. Amostras

Vamos considerar agora que o resultado de cada dado (1 a 6) como uma variável contínua (e não categórica, como tratamos acima). Dessa forma, o lançamento de cada dado é uma observação de uma amostra de tamanho n .

Para demonstrar o funcionamento do teorema do limite central, vamos calcular a média dos lançamentos de 10.000 amostras (de tamanho $n = 100$), e investigar a distribuição desse conjunto de médias amostrais.

```
set.seed(1234)

medias_amostrais <- NA # Cria o objeto em que serao guardadas as medias amostrais
sd_amostrais <- NA

n_amostra <- 100

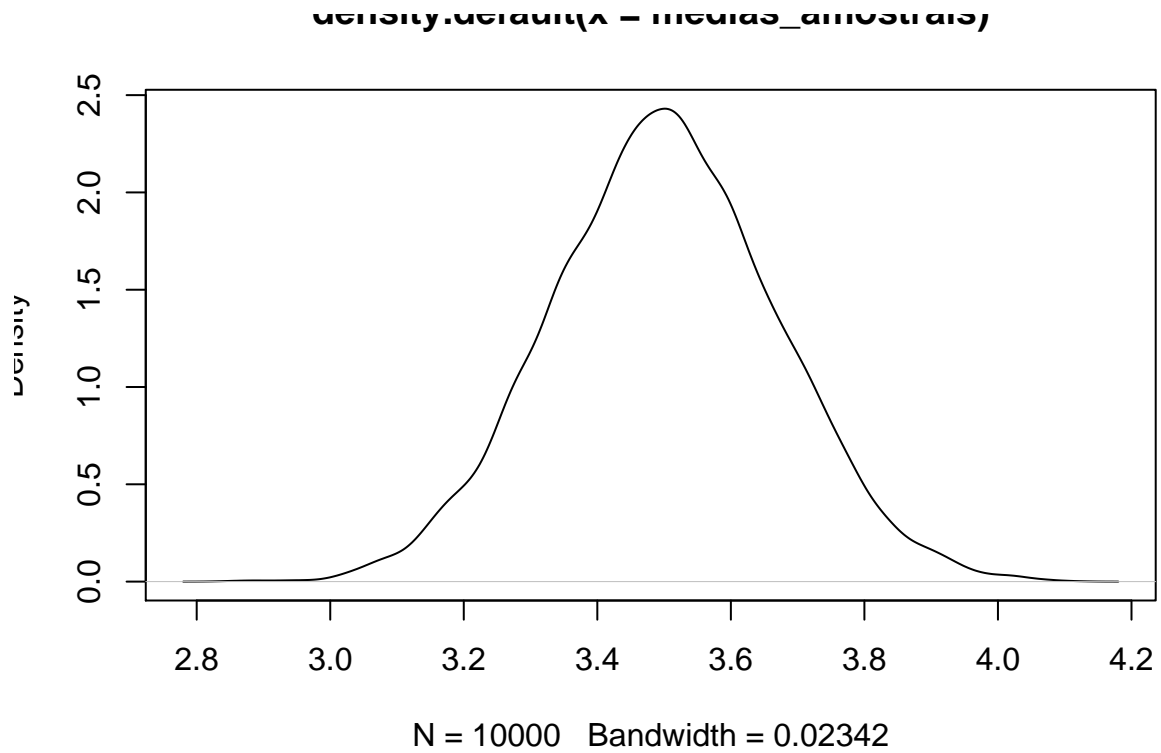
for(i in 1:10000){

  lancamento_dados <- sample(dado, size=n_amostra, replace=T)
  medias_amostrais[i] <- mean(lancamento_dados)
  sd_amostrais[i] <- sd(lancamento_dados)

}
```

Agora, nosso objeto `medias_amostrais` é um vetor com as 10.000 médias amostrais extraídas das amostras de tamanho $n = 100$. Um gráfico de densidade de kernel dessas médias amostrais nos revela uma distribuição bastante próxima à de uma distribuição normal:

```
plot(density(medias_amostrais))
```



E como se compara a média dos valores de um dado com a média das nossas médias amostrais?

```
mean(medias_amostrais)
```

```
[1] 3.499525
```

```
mean(dado)
```

```
[1] 3.5
```

Por fim, calcule o erro-padrão da nossa média amostral. Como ele se compara com o desvio-padrão das médias amostrais?

```
mean(sd_amostrais)/sqrt(n_amostra)
```

```
[1] 0.1706158
```

```
sd(medias_amostrais)
```

```
[1] 0.1693405
```