

## Lição 8: Conceitos Básicos de Estatística Inferencial

Nesta lição, vamos tratar de alguns conceitos centrais da Estatística Inferencial: *hipótese alternativa e hipótese nula; distribuição normal; significância; erros do Tipo I e do Tipo II; e intervalo de confiança.*

Até agora, fizemos tabelas e gráficos de dados de variáveis nominais, ordinais e numéricas. Reproduzo um desses gráficos na Figura 8.1: a proporção de apagamento de /r/ pós-vocálico em lojas de departamento de Nova Iorque na década de 1960, com base nos dados de Labov (1972). Vimos que houve menos r0 entre os funcionários da Saks, um pouco mais na Macy's, e mais ainda na S. Klein. Em nenhuma loja houve 0% ou 100% da variante “apagamento de /r/”.

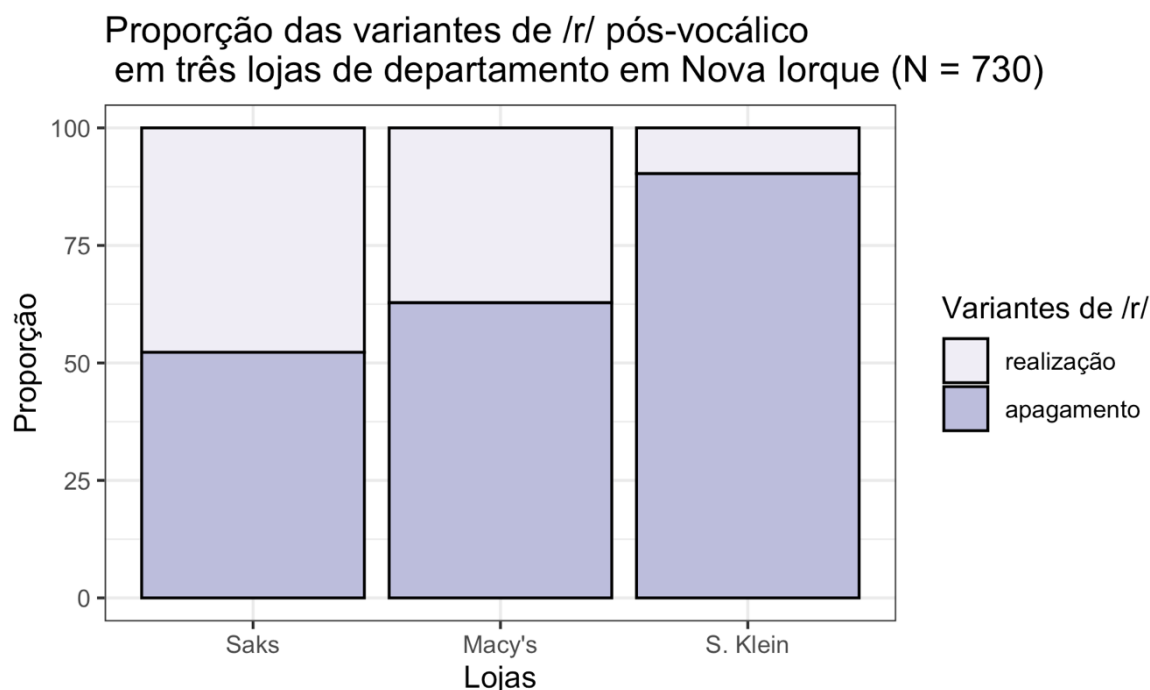


Figura 8.1: Gráfico de barras feito na Lição 5. Fonte: própria.

Na Figura 8.2 está outro gráfico que plotamos anteriormente: a distribuição das vogais pretônicas nas amostras de falantes paraibanos e paulistanos residentes em São

Paulo. Vemos que há diferenças entre as vogais e entre os dois grupos de falantes, mas que também há muita sobreposição no intervalo de F1 para cada vogal. Há dados, por exemplo, de /o/ cujas medidas de F1 são tão baixas quanto as de vogais /a/.

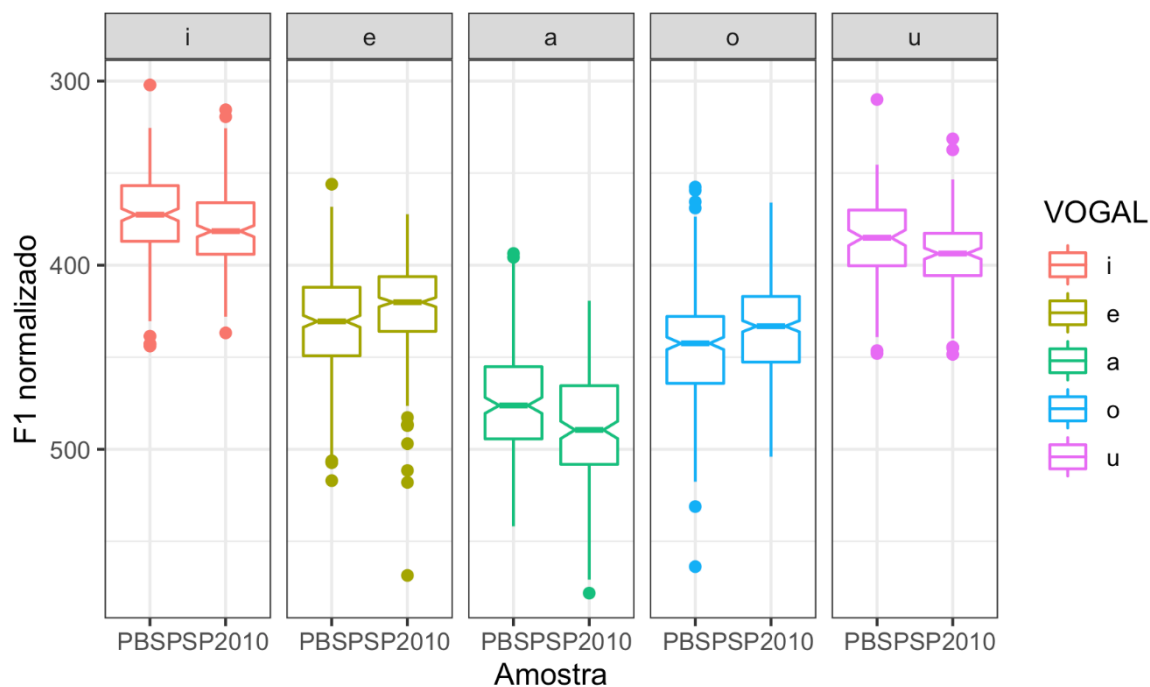


Figura 8.2: Boxplot feito na Lição 7. Fonte: própria.

Quando se lida com dados reais, inevitavelmente se depara com a variação. Há uma série de fenômenos no mundo – e na língua! – que não são categóricos. E, quando se trata de variação, sempre há um grau de incerteza. Quem vai ganhar a próxima eleição? Será que hoje vai chover?

O fato de que há variação no mundo, contudo, não implica que tudo seja caótico, regido pelo puro acaso e aleatoriedade. Mesmo nos casos em que determinado resultado não seja 100% certo, podemos estimar *probabilidades*. As pesquisas de intenção de voto antes de uma eleição e a previsão diária do tempo são exemplos disso.

Nos dados de vogais pretônicas, vemos que as vogais médias /e/ e /o/ são mais baixas na fala de paraibanos do que na de paulistanos. Mas com tanta variação e sobreposição de medições, uma pergunta que se coloca é: será que isso aconteceu por acaso? Será que, em outro conjunto de dados, a distribuição de F1 dessas vogais poderia

ter sido igual, ou quiçá se observassem vogais mais baixas entre paulistanos? Uma primeira resposta para essas perguntas é: sim, isso é *possível*. Mas quão *provável* isso é?

Em pesquisas científicas, levantamos hipóteses a serem testadas. Para que hipóteses possam ser colocadas à prova, elas precisam se configurar como uma afirmação que se refere a mais do que um único evento e precisam ser falseáveis (Gries, 2019, p. 22). A primeira dessas características parece ser autoevidente. Se uma hipótese não é generalizável para além de uma única observação, ela é de pouco uso para ser testada.

A falseabilidade diz respeito à própria possibilidade de colocar a hipótese à prova e refutá-la. Por exemplo, a seguinte afirmação pode ser falseada: “Alunos universitários cotistas têm desempenho escolar tão bom quanto ou melhor do que de alunos universitários não cotistas”.

A hipótese acima é chamada de H1 ou hipótese alternativa. A ela, opõe-se a H0, ou hipótese nula. A hipótese nula é normalmente formulada a partir da negação de H1. Neste exemplo, H0 seria “Alunos universitários cotistas não têm desempenho escolar tão bom quanto ou melhor do que de alunos universitários não cotistas”.

H1 + H0 devem abarcar todos os resultados possíveis. Se algum cenário ou resultado não é previsto nem por H1 nem por H0, a hipótese está mal formulada. Em um levantamento sobre o desempenho escolar de cotistas e não cotistas, só há três resultados possíveis: (i) cotistas têm melhor desempenho do que não cotistas; (ii) cotistas têm desempenho tão bom quanto de não cotistas; (iii) cotistas têm desempenho pior do que não cotistas. Todos esses cenários estão cobertos em H1 + H0. H1 e H0 são formuladas como afirmações categóricas para que sejam falseáveis. Mas, como vimos, a realidade dos dados nem sempre (na verdade, quase nunca) é categórica.

Vejamos um exemplo cotidiano para ver H1 e H0 em ação. O exemplo a seguir é adaptado de Gries (2019, p.34ss). Imagine que decidimos jogar cara e coroa. Eu sou cara e você é coroa, ok? Para tornar as coisas mais interessantes, apostamos dinheiro: toda vez que cair cara, você me dá 5 reais; toda vez que cair coroa, eu te dou 5 reais. Estabelecemos de antemão que a moeda será jogada 100 vezes.

Digamos que, ao final de 100 jogadas, eu tenha ganhado 52 vezes e você 48. Você diria que eu roubei? E se eu tivesse ganhado 54 vezes e você 46? E se eu tivesse ganhado 60 vezes e você 40?

Não há um limite absolutamente exato para decidir que alguém está roubando num jogo da moeda. Mas o ponto aqui é: a depender do resultado, você ia desconfiar seriamente de que eu roubei. Você nem coloca em questão o cenário oposto – você estar roubando, pois você, que é honesto, sabe que não está roubando; a dúvida é apenas sobre seu oponente. Se eu tivesse ganhado 90 vezes e você apenas 10, acho muito difícil que você pensasse: “Puxa, como ela é sortuda!...”. E o motivo disso é simples: embora 90 vs. 10 seja um resultado *possível*, ele é muito pouco *provável*.

Pare um pouco para pensar a partir de qual resultado você acharia que estou roubando (55? 60? 62?...). Digite esse número no *script*, atribuindo-o a uma variável chamada *x*.

```
x <- 60
```

**N.B.:** No exemplo acima está o número “60”, mas você pode escolher qualquer outro entre 51 e 100!

Você está efetivamente estabelecendo uma hipótese: minha oponente está roubando no jogo de moeda! Esta é a hipótese alternativa,  $H_1$ . A hipótese nula afirma o contrário: minha oponente *não* está roubando.

A beleza na Estatística está no desenvolvimento de métodos para calcular essas probabilidades. Vamos simplificar nosso jogo da moeda de 100 para 3 jogadas para facilitar a demonstração. Em três jogadas de moeda, há no total 8 resultados possíveis, discriminados na Tabela 8.1.

*Tabela 8.1: Resultados possíveis para três jogadas de cara ou coroa e suas respectivas probabilidades ( $H_0$  está correta).*

Jogada 1	Jogada 2	Jogada 3	# cara	# coroa	$p_{\text{resultado}}$
cara	cara	cara	3	0	0,125
cara	cara	coroa	2	1	0,125
cara	coroa	cara	2	1	0,125

cara	coroa	coroa	1	2	0,125
coroa	cara	cara	2	1	0,125
coroa	cara	coroa	1	2	0,125
coroa	coroa	cara	1	2	0,125
coroa	coroa	coroa	0	3	0,125

*Fonte: Gries (2019, p.37).*

Considerando que a moeda é honesta, não viciada, a probabilidade de cada um dos 8 resultados ocorrer é  $1 / 8 = 0,125$ . Com isso é fácil computar a probabilidade de cada resultado final: a probabilidade de 3 caras e 0 coroas (ou de 0 caras e 3 coroas) é 0,125; a probabilidade de 2 caras e 1 coroa (ou 1 cara e 2 coroas) é  $3 * 0,125 = 0,375$ .

É fácil calcular à mão quais são todos os resultados possíveis de três jogadas de moeda, mas ficaria mais difícil fazer isso em um número maior de jogadas. Para 100 jogadas, o total de possíveis resultados é  $2 ^ 100$  (em que  $^$  representa “elevado a”). Faça esse cálculo no R agora.

```
2 ^ 100
```

```
## [1] 1.267651e+30
```

Esse é um número com tantos dígitos que o R fornece o resultado em *notação e*: 1.267651e+30, que equivale a  $1,267651 * (10^{30})$ . Já é difícil fazer sentido desse número – quem dirá calcular as probabilidades de cada um desses possíveis resultados! Mas é claro que o R tem uma função para fazer esse cálculo.

Os resultados de um jogo de cara ou coroa seguem uma distribuição binomial, pois há dois resultados possíveis. Uma função para calcular probabilidades dentro de uma distribuição binomial é `dbinom()`. Em uma forma simples, ela toma como argumentos (i) o número de sucessos de uma variável binária (p.ex., 3 caras); (ii) o número de tentativas; (iii) a probabilidade de se ter aquele resultado em uma tentativa (para a moeda, 50%, ou 0,5). Digite então `dbinom(3, 3, 0.5)` para ver o resultado.

```
dbinom(3, 3, 0.5)
```

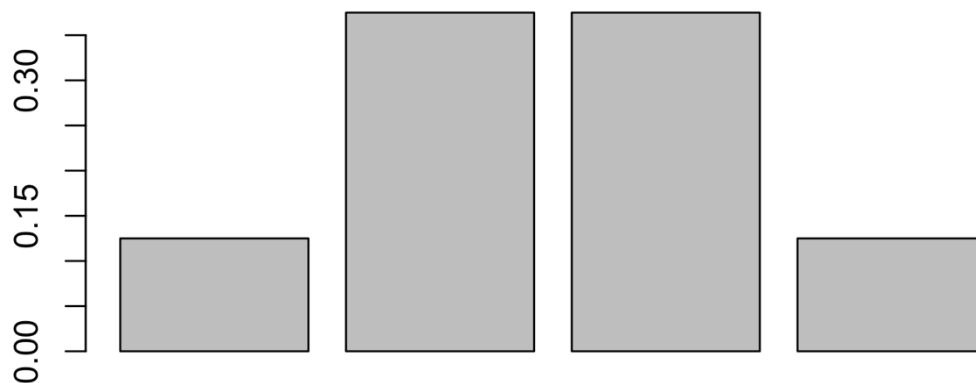
```
## [1] 0.125
```

O R informa que a probabilidade de se ter 3 caras em 3 jogadas de moeda, em que a probabilidade de cara é 0.5, é 0,125 – exatamente como calculamos acima. Podemos também pedir que o R calcule a probabilidade de todos os resultados possíveis – de 0 a 3 caras (o que também cobre os resultados de se ter 3 a 0 coroas). Digite `dbinom(0:3, 3, 0.5)` para ver o resultado.

```
dbinom(0:3, 3, 0.5)
## [1] 0.125 0.375 0.375 0.125
```

O R mostra, em ordem, qual é a probabilidade de se ter 0 cara, 1 cara, 2 caras e 3 caras. Que tal colocar isso num gráfico? Já vimos que é mais fácil compreender elementos gráficos do que textuais. Vamos usar aqui a função `barplot()`, da instalação base do R, para você também conhecê-la (e ver que o resultado é bem mais feinho que os gráficos do `ggplot2`, mas ele já vai dar conta do recado). Como primeiro argumento, use a função `dbinom()` do modo como especificamos acima. Como segundo argumento, use `beside = T`, para que as barras fiquem lado a lado. O resultado se encontra na Figura 8.3.

```
barplot(dbinom(0:3, 3, 0.5), beside = T)
```

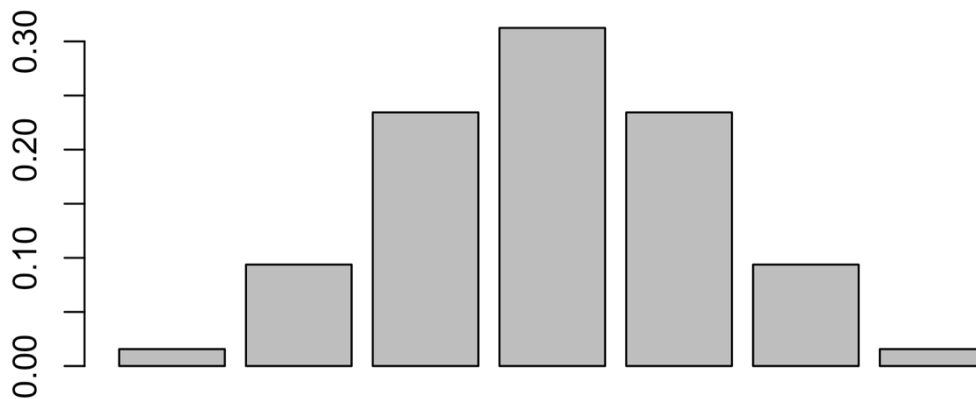


*Figura 8.3: Distribuição binomial da probabilidade de se obter 0 a 3 caras em três jogadas de moeda. Fonte: própria.*

As barras representam, da esquerda para a direita, os valores 0.125, 0.375, 0.375 e 0.125 computados acima para a probabilidade de 0, 1, 2, e 3 caras respectivamente.

Vamos agora visualizar quais são as probabilidades de se ter de 0 a 6 caras em 6 jogadas de moeda. A partir da linha de comando acima, substitua os dois '3' por '6' para obter `barplot(dbinom(0:6, 6, 0.5), beside = T)`.

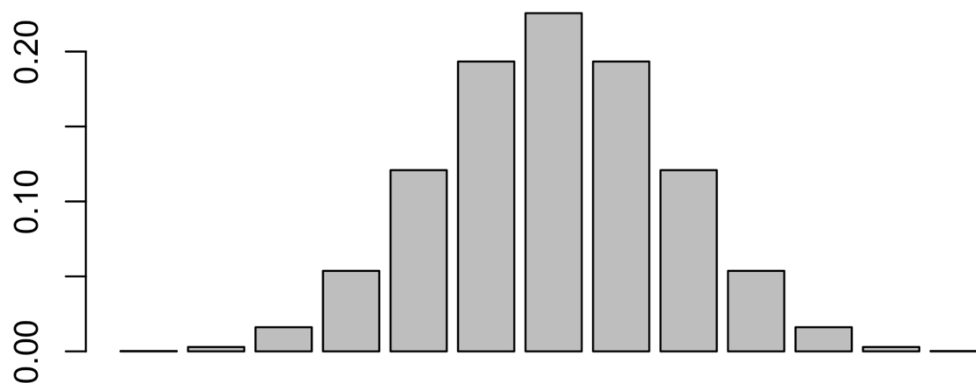
```
barplot(dbinom(0:6, 6, 0.5), beside = T)
```



*Figura 8.4: Distribuição binomial da probabilidade de se obter 0 a 6 caras em seis jogadas de moeda. Fonte: própria.*

Faça o mesmo para 12 jogadas de moeda. A partir da linha de comando acima, substitua '6' por '12'.

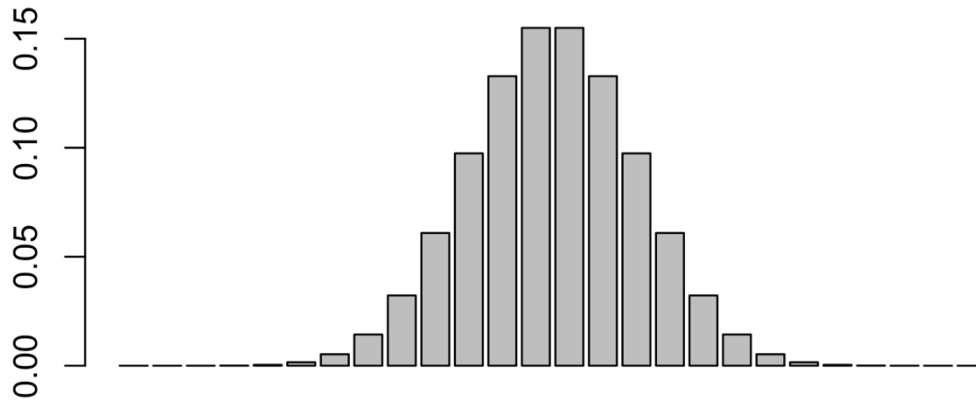
```
barplot(dbinom(0:12, 12, 0.5), beside = T)
```



*Figura 8.5: Distribuição binomial da probabilidade de se obter 0 a 12 caras em 12 jogadas de moeda. Fonte: própria.*

Faça o mesmo para 25 jogadas de moeda para computar todas as probabilidades de 0 a 25 caras.

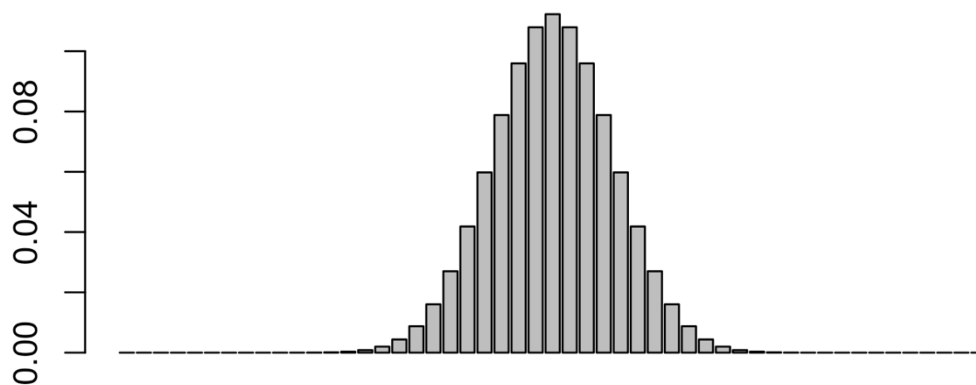
```
barplot(dbinom(0:25, 25, 0.5), beside = T)
```



*Figura 8.6: Distribuição binomial da probabilidade de se obter 0 a 25 caras em 25 jogadas de moeda. Fonte: própria.*

Faça o mesmo para 50 jogadas de moeda para computar todas as probabilidades de 0 a 50 caras.

```
barplot(dbinom(0:50, 50, 0.5), beside = T)
```



*Figura 8.7: Distribuição binomial da probabilidade de se obter 0 a 50 caras em 50 jogadas de moeda. Fonte: própria.*

Faça o mesmo para 100 jogadas de moeda para computar todas as probabilidades de 0 a 100 caras.

```
barplot(dbinom(0:100, 100, 0.5), beside = T)
```



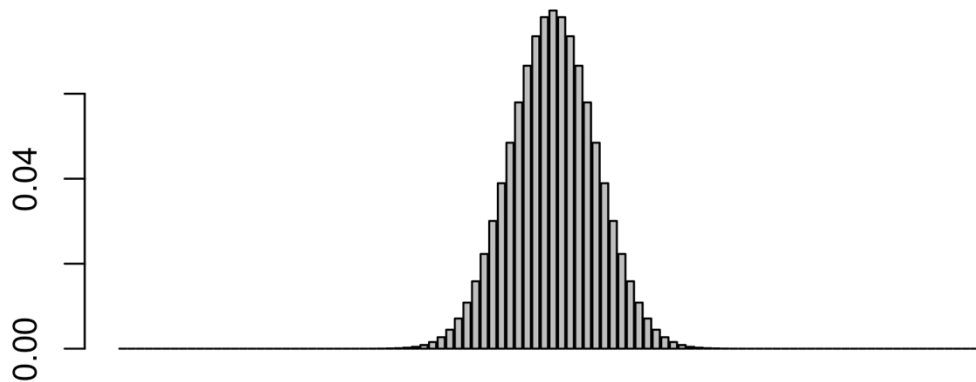


Figura 8.8: Distribuição binomial da probabilidade de se obter 0 a 100 caras em 100 jogadas de moeda. Fonte: própria.

Faça o mesmo para 1000 jogadas de moeda para computar todas as probabilidades de 0 a 1000 caras.

```
barplot(dbinom(0:1000, 1000, 0.5), beside = T)
```

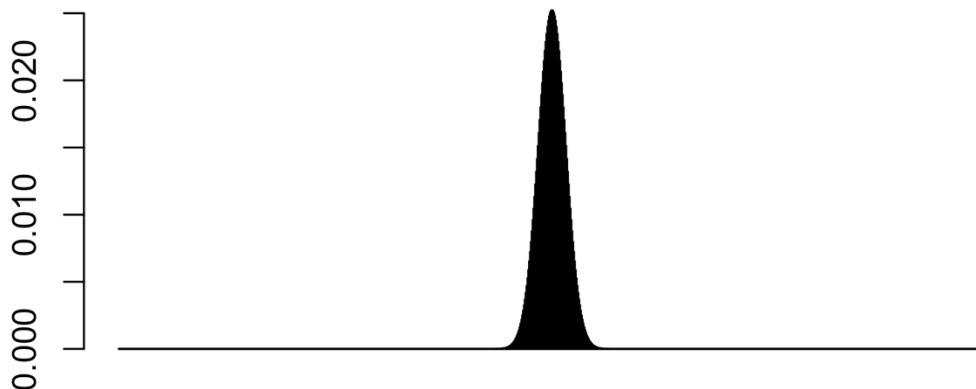


Figura 8.9: Distribuição binomial da probabilidade de se obter 0 a 1000 caras em 1000 jogadas de moeda. Fonte: própria.

Ok! Acho que deu pra pegar a ideia. Esses gráficos te lembram alguma coisa?

Sim! Trata-se da distribuição normal, que vimos em lições anteriores. Reveja os gráficos plotados na aba Plots por meio das flechas pra esquerda e pra direita. À medida que acrescentamos mais tentativas (3, 6, 12, 25 etc.), o gráfico de probabilidades se aproxima cada vez mais da distribuição normal, também conhecida como curva em forma de sino ou curva gaussiana.

Veja o eixo y da figura em que plotamos a probabilidade de 0 a 1000 caras em 1000 jogadas de moeda. A maior probabilidade – que é resultado de 500 caras em 1000 jogadas – está em torno de 0,025 (2,5%). Neste ponto, cada uma das probabilidades é pequena.

Não importa qual é o número N de tentativas, a soma de todas as probabilidades 0:N é 1. Faça essa checagem: aplique a função `sum()` ao cálculo das probabilidades de 0 a 3 caras em 3 tentativas. Digite `sum(dbinom(0:3, 3, 0.5))`.

```
sum(dbinom(0:3, 3, 0.5))
```

```
## [1] 1
```

Faça o mesmo agora para o cálculo de probabilidades de 0 a 100 caras em 100 tentativas. A partir da linha de comando acima, teste `sum(dbinom(0:100, 100, 0.5))`.

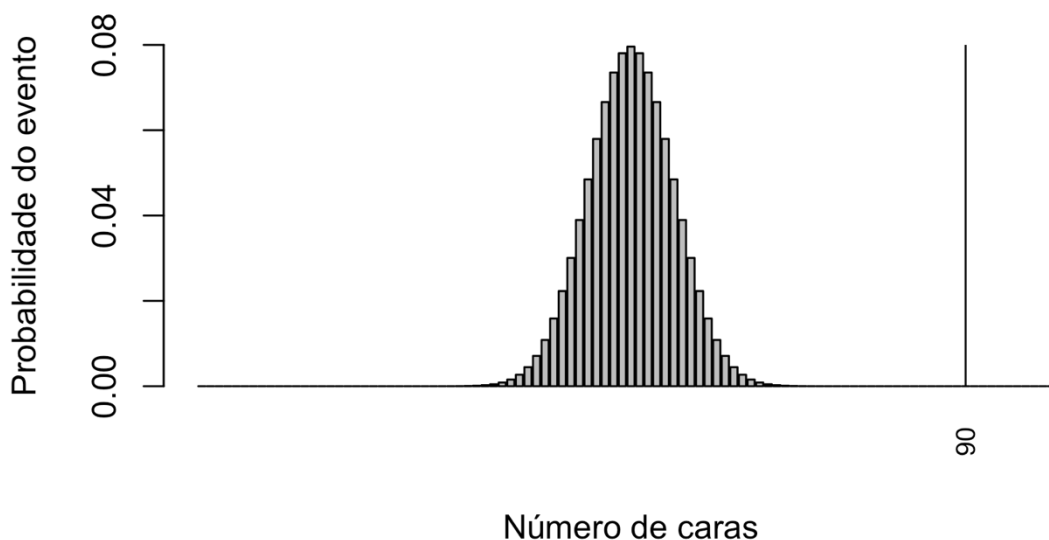
```
sum(dbinom(0:100, 100, 0.5))
```

```
## [1] 1
```

Ambos os resultados são 1. E é claro que tem que ser! 1 equivale a 100%, e é certeza de que um dos resultados vai ser um dos resultados possíveis. Isso te parece uma obviedade? Acompanhe o raciocínio, porque isso vai ter consequências importantes para o cálculo de probabilidades.

Podemos então estipular que a área total da figura que se forma pela distribuição binomial é 1 (ou 100%). Vemos também que as menores probabilidades se encontram nos extremos da distribuição, também chamados de “caudas”. Você consegue imaginar que o resultado de ter 90 caras em 100 jogadas de moeda está na cauda superior da distribuição, certo?

Imagine então se fizemos um corte vertical a partir da probabilidade de ter 90 caras. Exemplifico isso na Figura 8.10. Neste ponto, as probabilidades são tão pequenas que parecem ser zero – mas não são! É possível calcular a probabilidade de ter 90 caras ou mais, em 100 jogadas, fazendo o cálculo da área que vai de 90 até 100, concorda? Isso nada mais é do que a soma de todas as probabilidades a partir daquele ponto.



*Figura 8.10: Distribuição binomial da probabilidade de se obter 0 a 100 caras em 100 jogadas de moeda, com corte em 90 ou mais caras. Fonte: própria.*

Calculemos isso no R. A partir da linha de comando acima, em vez de colocar 0:100 no primeiro argumento, vamos colocar 90:100, para obter `sum(dbinom(90:100, 100, 0.5))`.

```
sum(dbinom(90:100, 100, 0.5))
```

```
## [1] 1.531645e-17
```

A probabilidade de ter qualquer resultado entre 90 e 100 caras, em 100 jogadas de moeda, é  $1.531645e-17$ , que o R novamente nos dá em notação e. Veja que, neste caso, o valor após ‘e’ (‘elevado a’) é negativo. Esse número equivale a  $1,531645 \cdot (10^{-17})$ . Quando elevamos um número a uma potência negativa, ele se torna um número menor do que 1. Em termos práticos, esse número é 0,0000... com 16 zeros após a vírgula e terminado em 1531645! Você pode pensar assim: o número negativo é o número de zeros, incluindo aquele antes da vírgula. Ou você ainda pode pensar: esse é um número absurdamente pequeno!!! (Veja também: <http://www.miniwebtool.com/scientific-notation-to-decimal-converter/>)

Ou seja, não só a probabilidade de ter 90 caras em 100 jogadas é pequena, mas também a soma de todas as probabilidades entre 90 e 100 é infinitesimalmente pequena. É esse conhecimento intuitivo que te faz desconfiar que um resultado de 90 caras em 100

jogadas, caso ocorresse, provavelmente não é fruto do acaso, e que te faz concluir, portanto, que sua oponente deve estar roubando! A probabilidade de isso acontecer *por puro acaso* é tão pequena que você é levado a rejeitar a hipótese nula (“minha oponente não está roubando”) e aceitar a hipótese alternativa (“minha oponente está roubando”). Essa probabilidade calculada é chamada de *valor-p*, ou *valor de significância*.

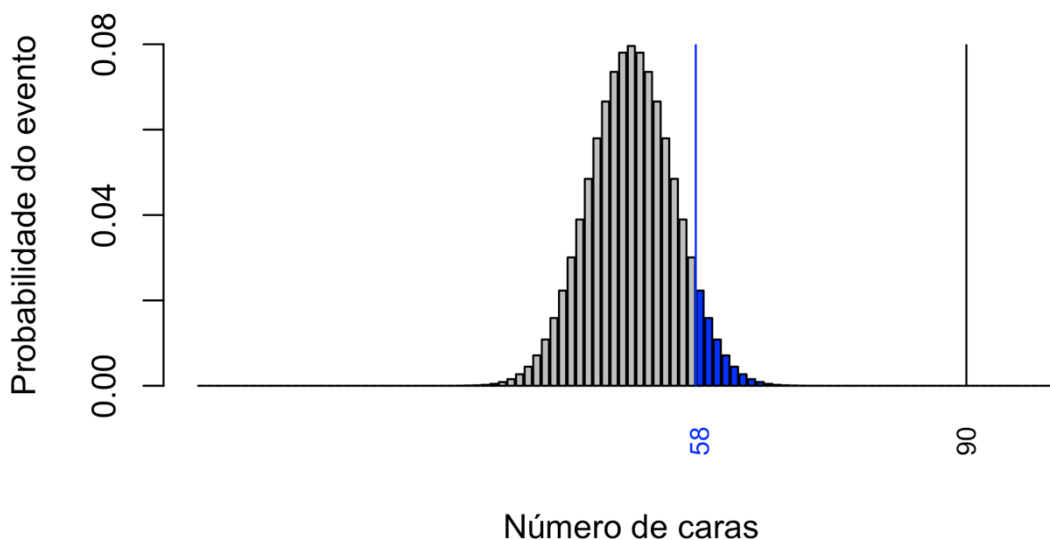
Uma probabilidade de  $1.531645e-17$  é tão pequena que logo a descartamos como extremamente improvável. Façamos agora o raciocínio inverso: a partir de qual probabilidade você consideraria algo muito pouco provável? A partir de 0,01%? 0,1%? 1%? 2%?

Não há um número certo aí. Mas, convencionalmente, a comunidade científica costuma usar o limite de 5% para considerar algo como muito pouco provável para acontecer ao acaso. Isso é chamado de nível  $\alpha$ : o limite estabelecido pelo pesquisador para rejeitar a hipótese nula. Note que, se ele é estabelecido pelo pesquisador, ele não necessariamente precisa ser 5%; você pode estabelecer um limite mais baixo, i.e., um teste mais rigoroso, antes de rejeitar a hipótese nula. O importante é respeitar o limite que você impôs, não mudá-lo de acordo com os resultados!

É possível também fazer o cálculo inverso ao que fizemos acima: se eu estabelecer um nível  $\alpha$  5%, a partir de qual resultado eu devo rejeitar  $H_0$  de que minha oponente não está roubando e aceitar  $H_1$ ? No R, isso pode ser calculado com a função `qbinom()`. Digite

```
qbinom(0.05, 100, 0.5, lower.tail = F)
qbinom(0.05, 100, 0.5, lower.tail = F)
## [1] 58
```

Na linha de comando acima, o que inserimos foi: o nível  $\alpha$  estabelecido (0.05), o número de tentativas (100), a probabilidade de o evento ocorrer (0.5) e para ignorar a cauda inferior da distribuição (só estamos atentando à cauda superior). O resultado é 58. Isso significa que a soma de todas as probabilidades de ocorrerem 58 caras ou mais é 5% – ou seja, já é bem baixa *em caso de a hipótese nula ser verdadeira* (= de sua oponente não estar roubando). Tal ponto é marcado na Figura 8.11, por meio da linha azul.



*Figura 8.11: Distribuição binomial da probabilidade de se obter 0 a 100 caras em 100 jogadas de moeda, com corte em 58 e em 90 ou mais caras. Fonte: própria.*

Como esse número – 58 – se compara com o valor que você estabeleceu quando pedi para que estipulasse um valor a partir do qual você consideraria que estou roubando? Para saber o nível  $\alpha$  que você estabeleceu para este teste, faça a soma das probabilidades de se obter de  $x$  a 100 caras, em 100 jogadas, quando a probabilidade de se ter cara é 50%, com uso de `sum()` e `dbinom()`.

```
sum(dbinom(x:100, 100, 0.5))
```

```
## [1] 0.02844397
```

**N.B.:** O resultado acima é para  $x = 60$ . Seu resultado será diferente se tiver estipulado outro valor para  $x$ .

Se você estabeleceu um valor mais alto do que 58, você queria ter muita certeza de que eu não estava roubando antes de sair fazendo acusações! Em outras palavras, você estabeleceu um nível  $\alpha$  mais rigoroso, abaixo de 5%. Se você estabeleceu um valor mais baixo do que 58, o nível  $\alpha$  foi menos rigoroso, acima de 5%. Este ponto está ilustrado na Figura 8.12 com a linha vermelha (neste exemplo, em 60, que foi o valor estabelecido para  $x$  acima).

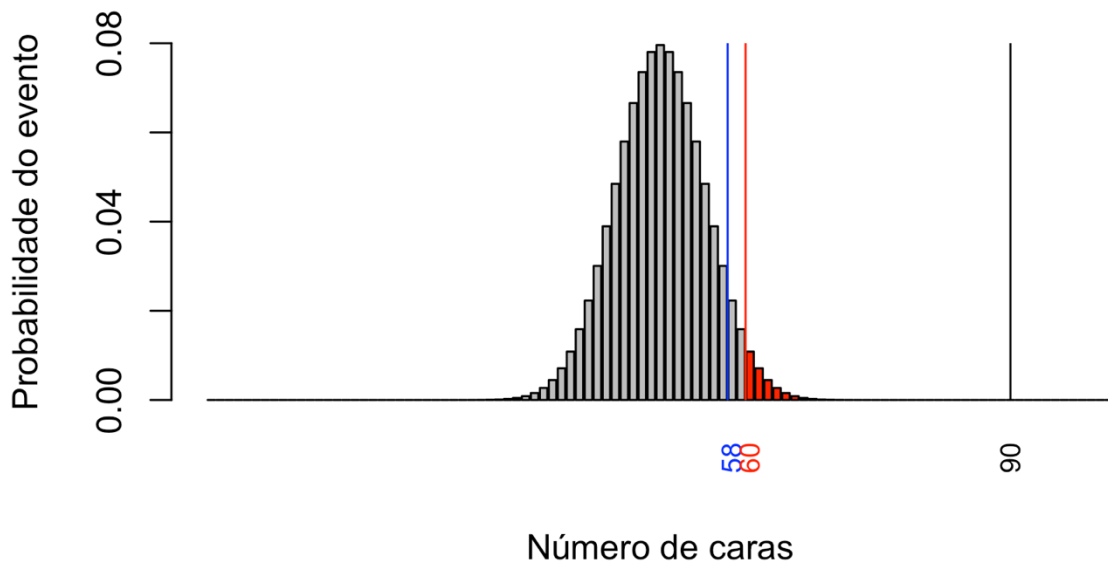


Figura 8.12: Distribuição binomial da probabilidade de se obter 0 a 100 caras em 100 jogadas de moeda, com corte em 58, 90 e  $x$  ou mais caras. Fonte: própria.

Neste exemplo, em que você é um dos jogadores e você *sabe* que não está roubando, sua hipótese só investigou a possibilidade de que *eu* pudesse estar roubando. Imagine agora outra situação: e se houver um juiz imparcial que está observando nossas jogadas? A partir de que ponto essa pessoa poderia começar a desconfiar de que um dos jogadores está roubando? Seria o mesmo limite (= 58 caras ou 58 coroas)?

- Sim
- Não
- Não faço ideia!

Pois é, o limite não seria o mesmo. Da perspectiva de uma terceira pessoa, é necessário olhar para ambos os lados da distribuição. Se o nível  $\alpha$  estabelecido permanecer 5%, há que se levar em conta 2,5% da cauda inferior e 2,5% da cauda superior (os resultados menos prováveis tanto de um lado quanto de outro).

Vamos fazer esse cálculo. A partir de uma das linhas de comando acima, em que usamos a função `qbinom()`, digite `qbinom(0.025, 100, 0.5, lower.tail = T)`, para verificar a cauda inferior.

```
qbinom(0.025, 100, 0.5, lower.tail = T)
```

```
## [1] 40
```

E digite `qbinom(0.025, 100, 0.5, lower.tail = F)`, para verificar a cauda superior.

```
qbinom(0.025, 100, 0.5, lower.tail = F)
```

```
## [1] 60
```

O R deu os resultados 40 e 60. No jogo da moeda, isso quer dizer que somente a partir de um resultado de 40 caras e 60 coroas (ou vice-versa) é que se poderia rejeitar a hipótese nula e assumir que um dos jogadores está roubando.

Este último é chamado de um *teste bidirecional*, e o teste anterior (em que você sabe que não está roubando) foi *unidirecional*. Ambos estão ilustrados na Figura 8.13. Note que, neste último tipo de teste (figura no topo), é mais fácil rejeitar  $H_0$ .

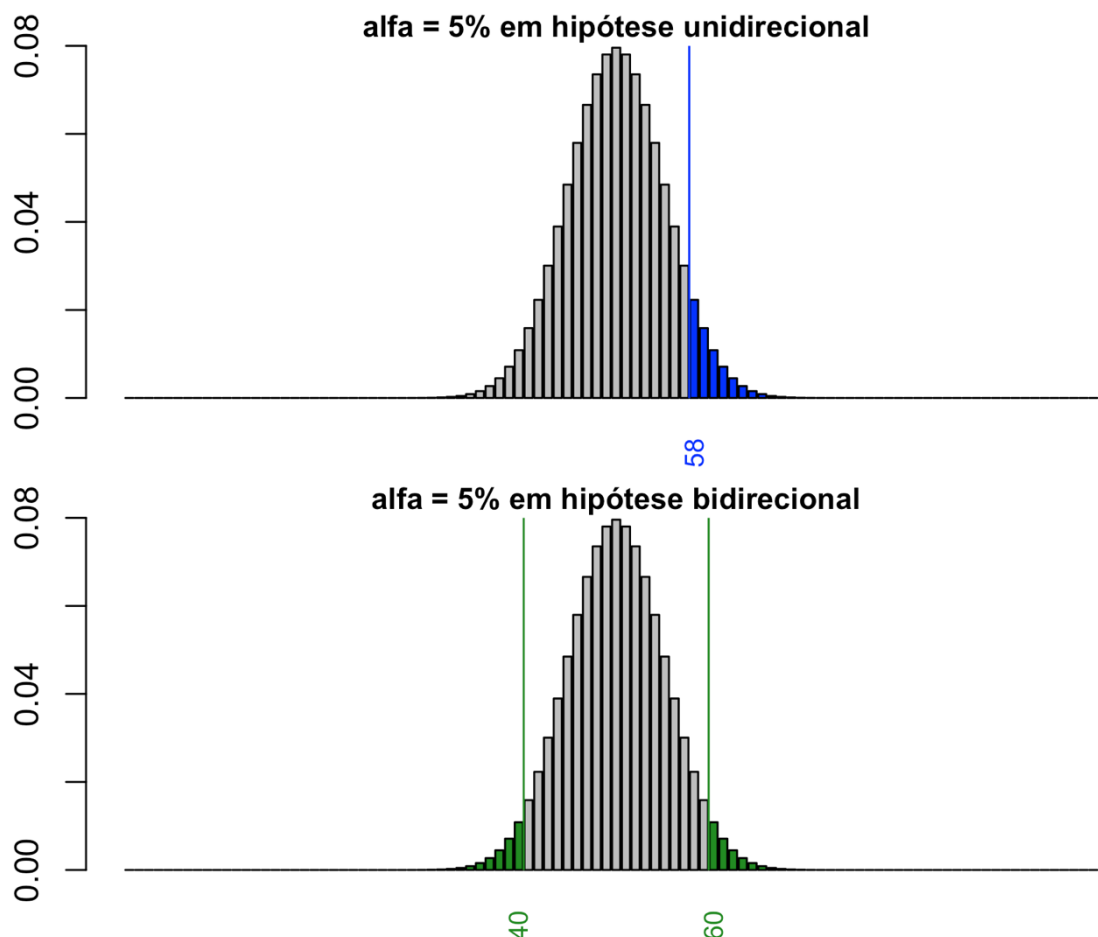


Figura 8.13: Distribuição binomial da probabilidade de se obter 0 a 100 caras em 100 jogadas de moeda, com corte em  $\alpha = 0,05$  em hipóteses uni e bidirecional.

Quando se aplica um teste uni ou bidirecional? No exemplo da moeda, você tinha o conhecimento prévio de que não estava roubando e, portanto, se alguém estava fazendo isso, só poderia ser eu. No caso de um juiz imparcial, ele não pode saber se qualquer dos jogadores está roubando, e precisa ficar atento a ambas as possibilidades.

A lição aqui é o fato de que conhecimento prévio compensa. Se você tem boas evidências – a partir da literatura sobre o assunto, por exemplo – de que determinado fenômeno ocorre de certa maneira, você pode rejeitar de antemão certos cenários.

Vamos pensar num exemplo linguístico. Para as vogais médias pretônicas /e/ e /o/ (em palavras como “relógio” e “romã”) – ver lições anteriores! –, temos, pela literatura, que paraibanos têm realização mais baixa (= F1 mais alto) do que a dos paulistanos. Neste caso, há boas evidências para que um pesquisador estabeleça, como H1, que os paraibanos residentes em São Paulo terão medidas de F1 mais altas do que para falantes paulistanos. Essa é uma hipótese unidirecional, que já estabelece de antemão a direção da diferença, se houver.

O que torna possível aplicar os mesmos cálculos no exemplo da moeda e em vogais pretônicas? É a distribuição normal! Você se lembra da forma da distribuição das medidas de F1 da vogal /e/ pretônica para paraibanos (Lição 7)? As mesmas propriedades aqui discutidas são válidas para diversos outros fenômenos variáveis no mundo.

Antes de prosseguir, veja o quadrinho da Figura 8.14.



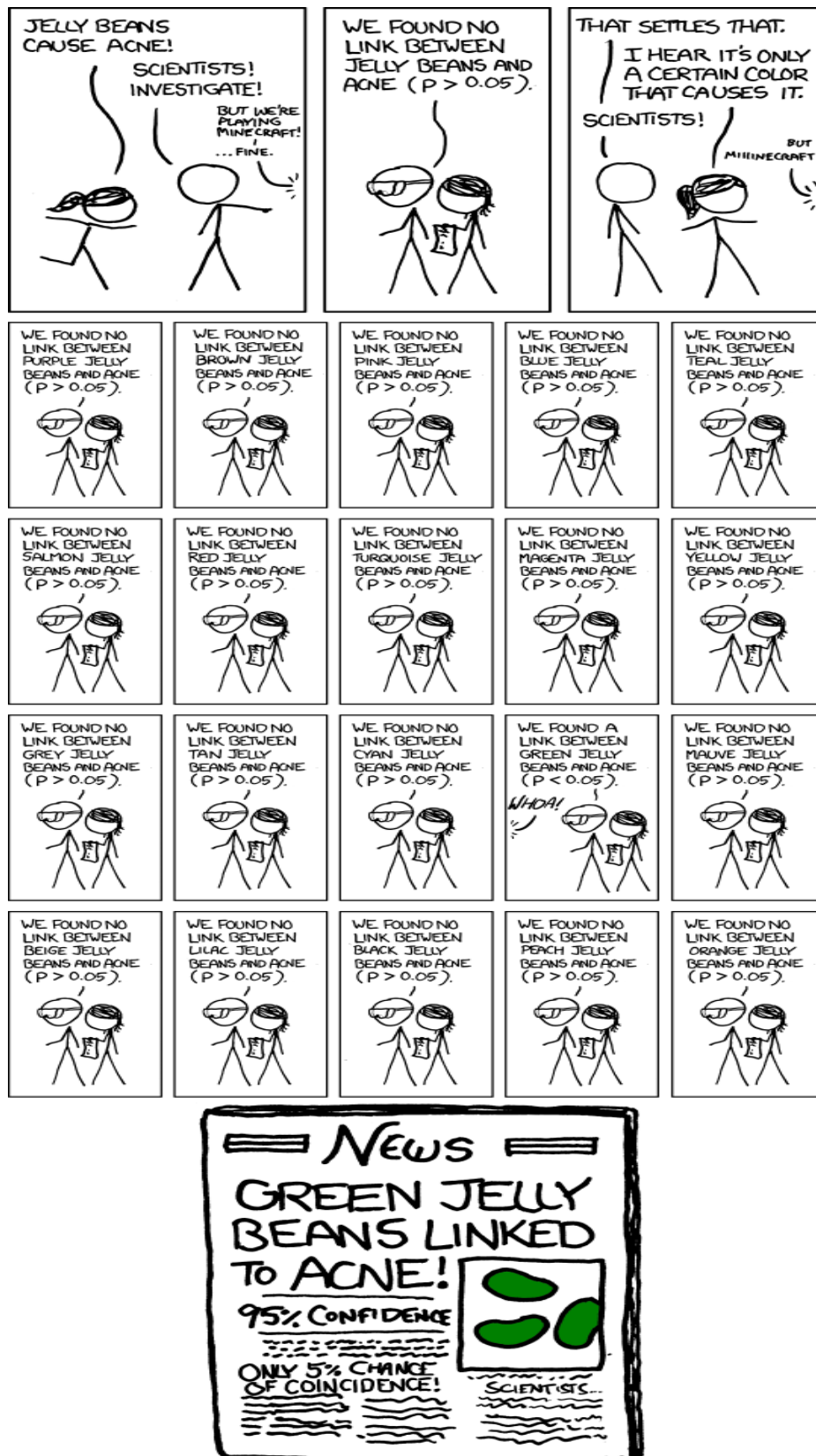


Figura 8.14: Significância. Fonte: <https://xkcd.com/882/>.

O quadrinho bem exemplifica o conceito de significância, assim como um problema para as análises estatísticas que decorrem do próprio conceito. Vimos que significância é a probabilidade de se observar determinada distribuição em caso de a hipótese nula ser verdadeira. Certas distribuições são tão improváveis que rejeitamos a hipótese nula e acatamos a hipótese alternativa.

Contudo, trata-se de uma medida de *probabilidade*. Isso significa que sempre há uma chance de que nossas conclusões foram errôneas. Se se estabelece um nível  $\alpha$  de 5%, existe uma probabilidade de que chegamos a conclusões equivocadas em média 5% das vezes, ou 1 em 20 – exatamente o que ocorre no quadrinho! O número de conclusões equivocadas tende a aumentar quanto mais testes realizamos – se realizamos 100 testes estatísticos, com nível  $\alpha = 5\%$ , as chances são de que 5 dos nossos resultados estão equivocados.

Um tipo de erro é o chamado *Erro do Tipo I*, que se refere à rejeição incorreta da hipótese nula. Trata-se de um falso positivo: encontramos um efeito que não existe de fato. Um problema é que não há como identificar qual ou quais dos resultados são falsos, uma vez que se trata de probabilidade.

Um problema correlato é fazer o oposto: erroneamente acatar a hipótese nula e não assumir a hipótese alternativa. Isso é chamado de *Erro do Tipo II*, que se refere a um falso negativo: não encontramos um efeito que existe de fato. Novamente, não temos como saber quais dos múltiplos testes aplicados geraram uma conclusão errônea.

O que seria um exemplo de Erro do Tipo I?

- numa sala de aula, não reprovar um aluno que colou na prova
- num tribunal, condenar um réu inocente
- as duas situações descritas

O que seria um exemplo de Erro do Tipo II?

- numa fábrica de automóveis, enviar para venda um carro com falha no airbag
- num laboratório, identificar como HIV positivo um paciente que não tem o vírus HIV

- as duas situações descritas

Se realizarmos 1000 testes com  $\alpha = 5\%$ , quantos falsos positivos esperamos obter em média?

- 5
- 50
- 500

Os erros tanto do Tipo I quanto do Tipo II podem ser graves. Felizmente, há métodos para contornar esse problema. Um desses métodos é a *Correção de Bonferroni*, cuja aplicação é bastante simples. Se se deseja realizar 20 testes com  $\alpha = 5\%$ , a correção de Bonferroni sugere que cada hipótese seja testada com  $\alpha 0,05 / 20 = 0,0025$ . Em outras palavras, o limite para se considerar um resultado significativo deve ser ajustado para que não se aumente a chance de uma conclusão errônea.

Outro método para diminuir a chance de falsos positivos ou falsos negativos é simplesmente não realizar todo e qualquer possível teste estatístico em determinado conjunto de dados. Uma pesquisa bem estruturada tem claras quais são as perguntas e as hipóteses que serão testadas, e não se perde em um sem número de questões.

Cabe ainda apresentar o conceito de *intervalo de confiança*. Nas últimas duas lições, calculamos as médias de valores de F1 e F2 das vogais pretônicas para plotar os espaços vocálicos de paraibanos e paulistanos. As médias resumem em um único número a estimativa de um parâmetro da população. Mas vimos também, pelos boxplots, que a distribuição real dos valores de formantes não se concentram em um único ponto, e sim são variáveis. A Figura 8.2 ilustra essa observação.

Existe a possibilidade, portanto, de que a média calculada esteja incorreta, e que seu verdadeiro valor esteja acima ou abaixo do valor estimado. O intervalo de confiança estabelece um valor mínimo e um valor máximo em que se calcula estar o verdadeiro parâmetro da população. Podemos não ter um valor exato, mas ter um intervalo em que é mais provável estar tal medida.

Quando se estabelece um nível  $\alpha$ , na prática, também se estabelece um *nível de confiança*. Quando  $\alpha = 5\%$ , o nível de confiança é de 95%; quando  $\alpha = 1\%$ , o nível de confiança é 99%. Nos boxplots, quando plotamos a figura com notch = T, o R faz o cálculo do intervalo de confiança e o representa graficamente pelos entalhes.

Por fim, é importante apontar para algumas concepções *errôneas* acerca do conceito de significância. Embora, no uso comum, dizer que algo é “significativo” implica dizer que é “importante” ou “relevante”, a significância estatística *nada* tem a ver com importância ou relevância. A significância ou valor- $p$  é uma medida de probabilidade de se observar determinada distribuição em caso de a hipótese nula ser verdadeira (no exemplo, a probabilidade de ter 90 caras em 100 jogadas se um dos jogadores não estiver roubando). Como tal, a medida não consegue avaliar quão importante é esse fato – só o pesquisador e a comunidade científica podem dizer.

Segundo, significância *não* é boa nem ruim. Ela simplesmente *é!* Um valor- $p$  abaixo de 5% *não* é intrinsecamente bom! Isso depende da hipótese levantada pelo pesquisador, da teoria, daquilo que se quer testar. Veja que, no estudo sobre vogais médias pretônicas na fala de migrantes, levanta-se a hipótese de que as vogais médias sejam mais baixas para os paraibanos do que para os paulistanos, mas há mais interesse em saber quem são os falantes que, em certa medida, não seguem essa regra – ou seja, aqueles que não apresentam diferenças significativas em relação aos paulistanos e para quem a hipótese nula não é rejeitada.

Por fim, um valor- $p$  ou de significância abaixo de 5% *não* prova que sua hipótese estava correta. O que se faz é *rejeitar a hipótese nula* e aceitar a hipótese alternativa, mas isso é diferente de dizer que H1 estava certa.

### **Para saber mais**

Recomendo a leitura de Gigerenzer (2004) para entender um pouco mais desses conceitos básicos e evitar a aplicação impensada de testes estatísticos.

**Exercícios**

1. Qual das afirmações a seguir é uma hipótese (H1) bem formulada? Explique sua resposta.
  - a. A medida média de F1 em Hz da vogal pretônica /e/ na fala de paraibanos é mais alta do que na fala de paulistanos.
  - b. O apagamento de /r/ pós-vocálico na fala de novaiorquinos possivelmente é mais frequente entre funcionários da Macy's do que os da Saks.
  - c. O próximo sintagma nominal plural a ser falado por mim será na forma padrão do português.
2. Qual é a H0 correspondente à H1 bem formulada acima?
  - a. A medida média de F1 em Hz da vogal pretônica /e/ na fala de paraibanos não é mais alta do que na fala de paulistanos.
  - b. O apagamento de /r/ pós-vocálico na fala de novaiorquinos possivelmente não é mais frequente entre funcionários da Macy's do que os da Saks
  - c. O próximo sintagma nominal plural a ser falado por mim não será na forma padrão do português.
3. Nesta lição, vimos o conceito de significância. Qual é o nível  $\alpha$  normalmente adotado como limite de significância?
  - a. 10%
  - b. 5%
  - c. 1%
  - d. 0,1%
  - e. 0,05%
4. Qual é a definição de significância?
  - a. a probabilidade de se observar determinada distribuição em caso de a hipótese nula ser verdadeira
  - b. a probabilidade de a hipótese nula ser verdadeira

- c. a possibilidade de que se chegou a uma conclusão errônea quando a hipótese nula é verdadeira
  - d. a possibilidade de rejeitar  $H_1$  equivocadamente em caso de a hipótese nula ser verdadeira
5. Numa rolada de dado, qual é a probabilidade de se ter o número 5, considerando-se que o dado não é viciado?
6. Numa rolada de dado, qual é a probabilidade de se ter um número ímpar, considerando-se que o dado não é viciado?
7. A probabilidade de obter três caras em três jogadas de moeda é  $1/8 = 0,125$ . Qual é a probabilidade de ter três números ímpares em três roladas de um dado, considerando-se que o dado não é viciado?
- a. 0,05
  - b. 0,125
  - c. 0,375
  - d. 0,50
8. Calcule a probabilidade de saírem 75 caras, em 100 jogadas de moeda, considerando-se que a moeda é honesta. Use a função `dbinom()` para seu cálculo.
9. Qual é o resultado da soma das probabilidades de todos os resultados possíveis em 1000 jogadas de moeda?
- a. 0,5
  - b. 1
  - c. 100
  - d. É impossível calcular
10. Calcule a soma das probabilidades de todos os resultados possíveis em 1000 jogadas de moeda. Para isso, use as funções `sum()` e `dbinom()`.
11. Qual é a notação decimal para o número em notação  $2e+05$ ?
12. Qual é a notação decimal para o número em notação  $1.45e-03$ ?
13. Qual não é outro nome para a forma da distribuição normal?

- a. curva em forma de Gauss
  - b. curva em forma de sino
  - c. distribuição gaussiana
14. Se se estabelece um nível  $\alpha = 1\%$  e se observa um valor de significância igual a  $2e^{-02}$ , o pesquisador deve...
- a. rejeitar a hipótese alternativa e acatar a hipótese nula
  - b. rejeitar a hipótese nula e acatar a hipótese alternativa
  - c. realizar o teste estatístico novamente
15. Quando um teste estatístico gera um  $p < 0,05, \dots$
- a. deve-se acatar a hipótese alternativa
  - b. tem-se confirmação de que a hipótese está correta
  - c. pode-se ter certeza da veracidade da hipótese alternativa
16. Rejeitar equivocadamente a hipótese nula é um...
- a. Erro do Tipo I
  - b. Erro do Tipo II
  - c. Erro do Tipo III
17. Erros do Tipo I e do Tipo II ocorrem...
- a. porque a medida de significância se refere a uma probabilidade e sempre há chance de erro
  - b. porque certos cuidados não foram tomados no momento da escolha do teste estatístico
  - c. quando não se aplica a correção de Bonferroni na realização de múltiplos testes estatísticos
18. O intervalo de confiança...
- a. estabelece um valor mínimo e um valor máximo dentre os quais há maior probabilidade de estar o verdadeiro parâmetro da população
  - b. estabelece o valor com maior probabilidade de ser o verdadeiro parâmetro da população

- c. estabelece um valor mínimo e um valor máximo do verdadeiro parâmetro da população
- d. estabelece o valor que tem 95% de probabilidade de ser o verdadeiro parâmetro da população