2

Testes Estatísticos

A leitura sistemática das revistas da área de saúde deixa evidente que os pesquisadores, embora trabalhem com *amostras*, *generalizam seus achados* para toda a *população* de onde a amostra foi retirada.[4] E, para generalizar seus achados, os pesquisadores aplicam *testes estatísticos*.

Vamos apresentar aqui a lógica desses testes por meio de um exemplo histórico. Em meados do século XIX, o cirurgião Joseph Lister (1827-1912), que estudava os trabalhos de Pasteur, percebeu que o estudo das bactérias e a prática da cirurgia eram ciências interdependentes.[5] Sendo isso verdade – Lister ponderou –, a assepsia das salas cirúrgicas deveria aumentar as taxas de sobrevivência nos pós-operatórios.

Para verificar essa hipótese, Lister fez um ensaio. Distribuiu 75 pacientes que iriam ser submetidos a uma cirurgia em dois grupos: o grupo controle, que foi submetido à cirurgia em salas nas condições usuais do hospital na época,[6] e o grupo com intervenção, que foi submetido à cirurgia em salas onde a assepsia havia sido feita com ácido fênico.

Lister tinha sua hipótese: a assepsia das salas cirúrgicas deveria aumentar a taxa de sobrevivência nos pós-operatórios. Outros cirurgiões tinham, porém, outra hipótese: a assepsia das salas cirúrgicas não teria qualquer efeito sobre as taxas de sobrevivência nos pós-operatórios. Os resultados do ensaio de Lister estão apresentados na Figura 2.1.

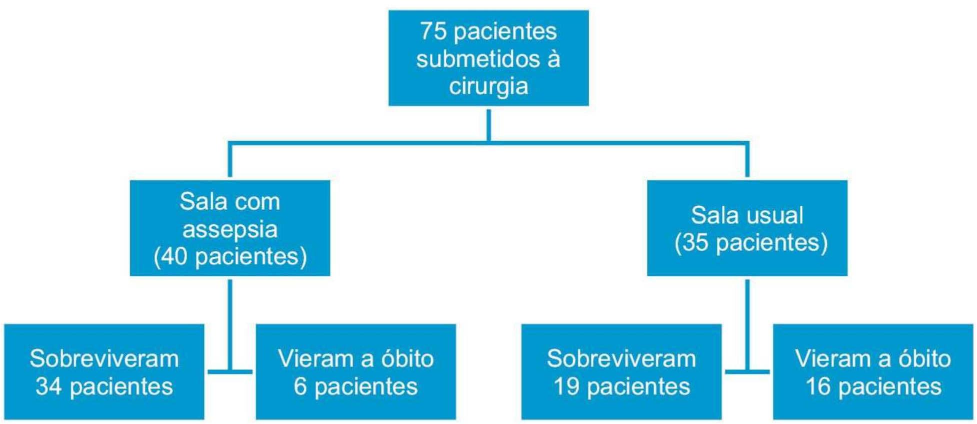


Figura 2.1 – Resultados do experimento de Lister.

A diferença das taxas de sobrevivência entre os dois grupos é grande. Veja: sobreviveram 34 dos 40 pacientes do grupo com intervenção. A taxa de sobrevivência é

$$\frac{34}{40} \times 100 = 85{,}0\%$$

Sobreviveram 19 dos 35 pacientes do grupo controle. A taxa de sobrevivência é

$$\frac{19}{35} \times 100 = 54{,}3\%$$

A diferença entre grupos é

$$85{,}0 - 54{,}3 = 30{,}7\%$$

Será que uma diferença entre grupos tão grande como essa foi suficiente para convencer os médicos da época da necessidade de assepsia? Não, como mostra a história da Medicina, mas como essa questão seria tratada hoje? Aplicando um teste estatístico.

2.1. DECIDINDO SOBRE AS HIPÓTESES

Vamos mostrar a construção de um teste estatístico usando o ensaio de Lister. Parece razoável considerar que há duas explicações possíveis para a diferença encontrada entre as taxas de sobrevivência:

1. *Acaso*. A diferença observada entre grupos seria obra do acaso.
2. *Assepsia*. A diferença observada entre grupos seria devida à intervenção (assepsia da sala cirúrgica).

Para decidir por uma das explicações, toda revista científica julga, hoje, mandatório aplicar um teste estatístico. Veja como se constrói o teste.

1. Comece considerando que a intervenção *não* tem efeito. Esta é a *hipótese da nulidade*[7] ("nulidade" porque diz que o efeito da intervenção é nulo). Indica-se por H_0 (lê-se agá zero). No exemplo, a hipótese é a de que a assepsia *não* tem efeito, ou seja, as taxas de sobrevivência *são iguais nos dois grupos*. Escreve-se:

   H_0: a intervenção *não tem* efeito.

2. Em seguida, considere a nova proposta: a intervenção *tem* efeito. Essa é a *hipótese alternativa.* Indica-se por H_1 (lê-se agá um). No exemplo, o pesquisador considerava que as taxas de sobrevivência *são maiores no grupo que recebeu a intervenção.* Escreve-se:

   H_1: a intervenção *tem* efeito positivo.

3. Agora, pergunte-se: se a *hipótese da nulidade for verdadeira,* é provável obter uma diferença entre taxas de sobrevivência tão grande quanto ou maior do que a que Lister obteve?
4. Calcule a probabilidade de ocorrer diferença entre taxas de sobrevivência tão grande ou maior do que a que Lister obteve, *considerando verdadeira a hipótese da nulidade*. Essa probabilidade[8] denomina-se *p-valor*.
5. Olhe o *p*-valor e admita:
   a) *p-valor pequeno* significa que é *pouco provável* obter uma *diferença tão grande ou maior* do que a observada se a *hipótese da nulidade for verdadeira.* Então, rejeite a hipótese da nulidade.
   b) *p-valor grande* significa que é *provável* obter *diferença tão grande ou maior* do que a observada se a *hipótese da nulidade for verdadeira.* Nesse caso, não rejeite a hipótese da nulidade.

Quão pequeno deve ser o *p*-valor para que você decida rejeitar a hipótese da nulidade? Por tradição (não há nenhuma razão teórica para isso), o *p*-valor é considerado pequeno quando:

a) É menor do que 0,05. Escreve-se $p < 0,05$. Diz-se, então, que o resultado é *significante* e a hipótese da nulidade é rejeitada. Veja a Figura 2.2. Ou, então,

b) É menor do que 0,01. Escreve-se $p < 0,01$. Diz-se que o resultado é *altamente significante*. A hipótese da nulidade é rejeitada.

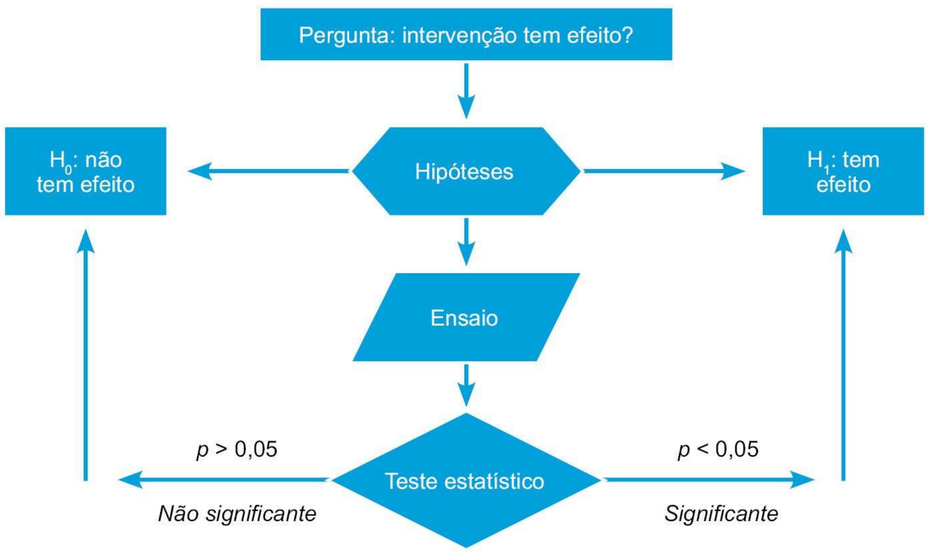


Figura 2.2 – Tomada de decisão: significante ou não significante?

Exemplo 2.1

Lister obteve os dados apresentados na Figura 2.1 . Sob a hipótese de que a assepsia não tem efeito sobre a taxa de sobrevivência dos operados (hipótese da nulidade), a probabilidade de ocorrer uma diferença igual ou maior do que a obtida por Lister é *p*-valor = 0,0024. Conclui-se então que a intervenção (assepsia) tem efeito.[9]

Sobrevivência de operados em salas cirúrgicas com e sem assepsia

Sobrevivência	**Assepsia na sala cirúrgica**		**Total**
	Sim	**Não**	
Sim	34	19	53
Não	6	16	22
Total	40	35	75
Taxa de sobrevivência	85,00%	54,30%	
p-valor	0,0024 < 0,05		

Fonte: WINSLOW, C. *The Conquest of Epidemic Disease*. Princeton: Princeton University Press, 1943. p. 303. Apud ALIAGA, M.; GUNDERSON, B. Interactive Statistics. 2nd ed. New Jersey: Prentice Hall, 2003. p. 673.

2.2. MEDINDO A INCERTEZA

2.2.1. Calculando o *p*-valor

Com base nos dados de uma amostra, os pesquisadores calculam o *p*-valor, que é a probabilidade de obter uma diferença tão grande ou maior do que a observada se o tratamento *não* tiver efeito.

Se o *p*-valor for pequeno, os pesquisadores fazem uma *inferência estatística:* concluem que o tratamento tem *efeito na população* de onde retiraram a amostra. No entanto, os pesquisadores não têm 100% de certeza de que essa conclusão está correta. Eles apenas sabem que a probabilidade de essa conclusão estar errada é muito pequena (é o *p*-valor).

Calcular o *p*-valor é difícil, mas hoje se usam computadores e programas prontos para fazer os cálculos. Antes da popularização dos computadores, usavam-se tabelas para determinar a significância dos resultados – e essas tabelas ainda são fornecidas nos livros de Estatística e na internet.

Não pense, porém, que ficou fácil aplicar um teste estatístico e obter o *p*-valor. É preciso, primeiramente, *escolher o teste* que será utilizado em função das hipóteses que se quer testar. Depois, é preciso verificar se a variável em análise atende às *pressuposições exigidas* para a aplicação do teste.[10]

E tenha muito cuidado na interpretação do *p*-valor,[11] que não é intuitiva porque usa a contradição: a hipótese colocada em teste afirma o contrário do que o pesquisador acha que é (ou gostaria que fosse) verdadeiro. Talvez por isso o *p*-valor seja, muitas vezes, mal interpretado. Para evitar que isso aconteça com você,[12] tenha sempre em mente que:

p-valor é a *probabilidade de obter uma diferença entre grupos tão grande ou maior do que a obtida* quando essa diferença não existe.

Em outras palavras:

p-valor é a probabilidade de o pesquisador estar errado quando diz que os grupos em comparação são diferentes.

Veja bem: o *p*-valor nada diz sobre o fato de H_0 *ser ou não ser verdadeiro*. Informa, apenas, que a decisão de rejeitar H_0 pode estar errada. No entanto, é o pesquisador quem deve resolver – juntando todas as informações que tem ao *p*-valor que obteve – se a hipótese da nulidade parece improvável o bastante para que ele corra o risco de dizer que *deve ser rejeitada*.

2.2.2. Apresentando o nível de significância

Para aplicar um teste estatístico, o pesquisador começa levantando a hipótese de que o tratamento *não* tem efeito. É a *hipótese da nulidade*, que se indica por H_0 (lê-se agá-zero). No entanto, o pesquisador acredita no contrário: acha que o novo tratamento tem efeito. Esta é a *hipótese alternativa* – que se indica por H_1 (lê-se agá-um).

O pesquisador faz o teste estatístico, que fornece o *p*-valor. Se o *p*-valor for pequeno – tradicionalmente, menor do que 0,05 –, o pesquisador *conclui que o tratamento tem efeito* na população de onde a amostra foi retirada. O pesquisador sabe que o *p*-valor é a probabilidade de essa conclusão estar errada.

Em teoria, o pesquisador deve estabelecer a probabilidade de errar quando diz que existe diferença entre grupos – e essa diferença não existe – antes de proceder ao teste. Essa probabilidade é o *nível de significância* do teste, que se indica por α (lê-se alfa). Quando você lê $p \leq 0{,}05$, saiba que a hipótese da nulidade (de que o tratamento não tem efeito) foi rejeitada no *nível de significância* de *5%*.

Teoricamente, a probabilidade de rejeitar H_0 quando H_0 é verdadeiro (o *nível de significância* α) deve ser estabelecida antes de o teste ser aplicado.

Exemplo 2.2

Lister considerou que a assepsia das salas cirúrgicas aumenta as taxas de sobrevivência nos pós-operatórios. Conduziu, então, um ensaio para mostrar isso. Hoje em dia, na publicação dessa pesquisa, obrigatoriamente estariam descritos:

- H_0: Na população, as taxas de sobrevivência são iguais nos dois grupos: tratado (com assepsia) e controle (sem assepsia).
- H_1: Na população, a taxa de sobrevivência no grupo tratado (com assepsia) é maior do que a taxa de sobrevivência no grupo controle (sem assepsia).
- Nível de significância: $\alpha = 0{,}05$.

Obtidos os dados, teria sido feito um teste estatístico para comparar as taxas de sobrevivência. Sob H_0, a probabilidade de ocorrer uma diferença igual ou maior do que a obtida por Lister é *p*-valor = 0,0024. Como o *p*-valor é menor do que o nível de significância adotado, isto é, $p < 0{,}05$, a conclusão é de que o ensaio trouxe evidência de que a assepsia nas salas de cirurgia aumenta a taxa de sobrevivência em cirurgias de amputação ($\alpha = 5\%$).

2.2.3. Avaliando o poder do teste

Poder[13] *do teste estatístico* é a probabilidade de o teste rejeitar a hipótese da nulidade quando a hipótese da nulidade é falsa.

Em outras palavras:

Poder do teste estatístico é a probabilidade de o teste detectar uma diferença que realmente existe na população.

Exemplo 2.3

Imagine que você está conduzindo uma série de ensaios com uma droga eficaz. Estão em comparação dois grupos: o tratado (que recebe a droga) e o controle (que não recebe a droga). Se você aplicar um teste estatístico com poder de 0,90 para comparar grupos, deve esperar resultados estatisticamente significantes em 90% das vezes. Em 10% dos casos, não deve esperar resultados estatisticamente significantes. Então, poder do teste é a probabilidade de encontrar a diferença entre os dois grupos quando essa probabilidade realmente existe.

Vários fatores afetam o poder de um teste estatístico. Vamos mencionar alguns, imaginando que você esteja aplicando sempre o *mesmo teste.*[14]

- O poder do teste depende do *tamanho da diferença*: os testes estatísticos têm mais poder para detectar uma diferença de 40% do que uma diferença de 2%.
- O poder do teste depende do *nível de significância adotado.* No *nível de significância* de *5%*, a hipótese da nulidade será rejeitada se $p < 0{,}05$. *Diminuir o nível de significância* faz diminuir a probabilidade de rejeitar a hipótese da nulidade – seja ou não verdadeira – e, consequentemente, *diminui o poder do teste.*
- O poder do teste depende do *tamanho da amostra.* A confiança na informação aumenta quando aumenta a quantidade de dados.

IMPORTANTE: Quando se calcula o tamanho da amostra, é comum adotar – embora não haja qualquer justificativa teórica para isso – nível de significância de 5% e poder de teste de 80%. Isso significa que você admite até 5% de probabilidade de estar errado ao dizer que os grupos são diferentes e 80% de probabilidade de detectar uma diferença que realmente existe.

2.3. TESTES UNILATERAIS E BILATERAIS

Para aplicar um teste de hipóteses é preciso, primeiramente, decidir se o teste será unilateral ou bilateral.

Um teste é *unilateral* quando a hipótese da *nulidade* só pode ser rejeitada se a diferença entre grupos tiver o sentido especificado pelo pesquisador (p. ex., o pesquisador diz, no projeto, que espera efeito positivo do novo tratamento).

Exemplo 2.4

Foi apresentado o resultado do teste estatístico para os dados de Lister. Veja as hipóteses:

- H_0: assepsia *não tem efeito* sobre a taxa de sobrevivência dos operados.
- H_1: assepsia *aumenta* a taxa de sobrevivência dos operados.

O teste unilateral se justifica: é razoável verificar se os dados apresentados por Lister já mostravam o que se sabe hoje.

Um teste é *bilateral* quando a hipótese da *nulidade* é rejeitada *qualquer que seja o sentido (o sinal) da diferença entre grupos*.

Exemplo 2.5

Dez ratos machos adultos, criados em laboratório, foram separados aleatoriamente em dois grupos: o grupo controle, que recebeu a ração normalmente usada no laboratório, e o grupo tratado, que recebeu uma ração experimental. Como não se sabe o sentido da diferença, isto é, se a nova ração vai aumentar ou diminuir o peso dos ratos, parece razoável fazer um teste bilateral:

- H_0: em média, o peso dos ratos tratados com a nova ração *é igual* ao peso dos ratos tratados com a ração conhecida.
- H_1: a média dos pesos dos ratos que receberam a nova ração é estatisticamente *diferente* da média dos ratos tratados com a ração conhecida.

Os estatísticos em geral discordam da ideia de testes unilaterais, que seriam válidos somente quando fosse *impossível* ocorrer uma diferença em sentido contrário ao esperado. Veja as razões por que eles discordam da aplicação de testes unilaterais:

1. Os testes bilaterais são *mais seguros* – porque o tratamento pode dar resultado contrário ao esperado.
2. Os testes bilaterais são *mais conservadores*, isto é, têm menor probabilidade de rejeitar H_0. Como grande parte dos experimentos não satisfaz às exigências metodológicas nem às pressuposições exigidas para aplicar os testes estatísticos, é melhor trabalhar com testes que têm menor probabilidade de detectar significância.

Alguns pesquisadores preferem o teste unilateral, que é referido como *teste de superioridade*. O argumento em favor dos testes unilaterais é o de que uma boa revisão da literatura mostra o sentido da diferença. No entanto, a situação será constrangedora se a diferença for significante, mas em sentido contrário ao esperado.

2.4. TESTES PARAMÉTRICOS E NÃO PARAMÉTRICOS

Na área médica, são muito usados o teste *t* de Student,[15] a análise de variância e o teste de Tukey.[16] Tais testes exigem, para sua aplicação, que a variável em análise seja *numérica* e as hipóteses sejam feitas sobre *parâmetros*, como médias e variâncias.[17] Daí o nome: *testes paramétricos*. Por exemplo, para comparar a altura de meninos e meninas da mesma idade, o pesquisador pode optar por testes paramétricos – e testar as diferenças de médias (pelo teste *t* de Student) e de variâncias (pelo teste *F* de Snedecor).

Os testes paramétricos têm, ainda, outras exigências. Por exemplo, o teste *t* exige pressupor que a variável em análise tenha distribuição normal ou, pelo menos, simétrica. O teste *F* exige, além da pressuposição de variável com distribuição normal ou aproximadamente normal, ou pelo menos simétrica, a pressuposição de homocedasticia (homogeneidade de variâncias, ou seja, a mesma variabilidade dentro de grupos).

E o que faz o pesquisador quando verifica que a variável que estuda não atende às pressuposições exigidas para procedimento de um teste paramétrico? Aplica um teste não paramétrico. Esses testes não exigem que a variável em análise seja numérica e não exigem pressuposição a respeito da distribuição da variável. *Não pense*, porém, que pode "substituir" um teste paramétrico por um não paramétrico: as hipóteses são diferentes. Quando você aplica um teste *t* de Student (paramétrico), a hipótese da nulidade é de que *as médias populacionais são iguais*. E quando você aplica um teste de Mann-Whitney (não paramétrico), a hipótese da nulidade é de *que as distribuições são iguais.*[18]

Exemplo 2.6

Imagine que um pesquisador quer verificar se a dipirona é mais eficaz do que o paracetamol no controle da dor após determinada intervenção odontológica. A intensidade da dor pode ser registrada por meio de notas, atribuindo valor 0 para nenhuma dor e valor 5 para dor intensa. Como a intensidade de dor é variável *ordinal, não tem sentido* calcular médias e, portanto, *não tem sentido* aplicar um teste paramétrico. O pesquisador deve buscar um teste não paramétrico.

A lógica dos testes não paramétricos é simples e aplicá-los é bem mais fácil do que aplicar seus equivalentes paramétricos. No entanto, como tudo na vida tem um preço, você não vai se surpreender com o fato de os testes não paramétricos terem algumas desvantagens:

- Os testes paramétricos são mais *poderosos*, ou seja, têm maior probabilidade de rejeitar a hipótese da nulidade quando essa hipótese é falsa.
- Os testes paramétricos são bastante *robustos*, ou seja, pequenas transgressões às pressuposições exigidas não invalidam os resultados – principalmente se as amostras forem de tamanho grande ou moderado.

Portanto, para a análise de dados de variáveis com distribuição aproximadamente normal obtidas de grandes amostras, a melhor opção ainda é um teste paramétrico, mais poderoso –

mesmo que algumas pressuposições não estejam totalmente satisfeitas. Se você optar por um teste não paramétrico, estará optando por um teste com *menos poder*.

De qualquer forma, diante de um problema real de pesquisa, você precisa se decidir por um teste paramétrico ou por um teste não paramétrico. Aplique um teste paramétrico se as seguintes afirmativas forem verdadeiras:

1. Dados quantitativos.
2. As pressuposições exigidas para a aplicação do teste escolhido estão satisfeitas ou são pequenas as transgressões.
3. Amostra é de tamanho grande ou, pelo menos, moderado.

Aplique um teste não paramétrico quando estiver diante de uma das seguintes situações:

1. Dados qualitativos.
2. As pressuposições exigidas para a aplicação de testes paramétricos não estão satisfeitas.
3. A amostra é pequena.
4. Existem dados discrepantes ou censurados, que podem tornar mais indicado calcular medianas – e não as médias.

2.4.1. Algumas indicações

Testes para comparar dois grupos independentes:

- Para a análise de variável numérica com distribuição normal ou, pelo menos, simétrica e amostra grande (maior do que 30), aplique um teste paramétrico: o teste *t* de Student.[19]
- Para a análise de variável ordinal ou de variável numérica com distribuição não simétrica e/ou amostra muito pequena, aplique teste não paramétrico: o teste de Mann-Whitney ou o teste da mediana, que serão tratados no Capítulo 3 deste livro.
- Para a análise de variável nominal, use o teste de χ^2 de Pearson ou o teste exato de Fisher, que serão tratados no Capítulo 5 deste livro.

Testes para comparar dois grupos dependentes:

- Para a análise de variável numérica com distribuição normal ou, pelo menos, simétrica e/ou uma amostra bastante grande, aplique um teste paramétrico: o teste *t* de Student para amostras pareadas.[20]
- Para a análise de variável ordinal ou de variável numérica com distribuição não simétrica e/ou amostra muito pequena, aplique um teste não paramétrico: o teste de Wilcoxon ou o teste do sinal, que serão tratados no Capítulo 4 deste livro.
- Para a análise de variável nominal, use o teste de χ^2 de Mc-Nemar,[21] tratado no Capítulo 5 deste livro.

Testes para comparar mais de dois grupos independentes:

- Se a variável for numérica e tiver distribuição normal ou aproximadamente normal, ou tiver pelo menos distribuição simétrica e a amostra for bastante grande, faça uma análise de variância[22].

- Se a variável for ordinal ou numérica, mas a distribuição não for simétrica e/ou a amostra for pequena, aplique um teste não paramétrico: o teste de Kruskal-Wallis ou o teste da mediana, mostrados no Capítulo 4 deste livro.

Testes para comparar mais de dois grupos dependentes:

- Se a variável for numérica e tiver distribuição normal ou aproximadamente normal, ou tiver pelo menos distribuição simétrica e a amostra for bastante grande, faça uma análise de variância.[23]
- Se a variável for ordinal ou numérica, mas a distribuição não for simétrica e/ou a amostra for pequena, aplique um teste não paramétrico: o teste de Friedman, que veremos no Capítulo 4 deste livro.

RESUMO E OBJETIVO DO CAPÍTULO

Neste capítulo foi apresentada a lógica de um teste estatístico. São definidos *p*-valor, testes unilaterais e bilaterais, testes paramétricos e não paramétricos. Portanto, uma vez lido este capítulo, você deverá saber o que significam:

1. Hipóteses estatísticas.
2. *p*-valor e nível de significância.
3. Poder do teste.
4. Teste unilateral.
5. Teste bilateral.
6. Testes paramétricos e não paramétricos.

---

[4] Para estudar amostragem, veja: VIEIRA, S. *Introdução à Bioestatística*. 5. ed. Rio de Janeiro: Elsevier, 2016.

[5] DE PAOLLO, C.; PASTEUR, Lister: a chronicle of scientific dependence. The Victorian Web: Literature, History and Culture in the age of Victoria. Disponível em: http://www.victorianweb.org/science/health/depaolo.html. Acesso em 23 de outubro de 2017.

[6] Meados do século XIX.

[7] A maioria dos estatísticos brasileiros diz "hipótese nula", traduzindo assim, de maneira a meu ver equivocada, a expressão inglesa *null hypothesis*.

[8] É difícil calcular essa probabilidade, mas esse cálculo é feito em computador.

[9] Foi feito um teste unilateral para comparar proporções.

[10] Nos próximos capítulos, veremos a indicação de vários testes e as fórmulas de cálculo. Usaremos tabelas disponíveis no final deste livro e mostraremos exemplos de resultados obtidos por meio de programas para computador.

[11] Esta argumentação está totalmente baseada em Motulsky, H. *Intuitive Statistics*. New York: Oxford University Press, 1995.

[12] Glantz, S. A. *Primer of Biostatistics*. 2nd ed. New York: McGraw, 1987. p. 97.

[13] Alguns autores dizem potência do teste (tradução de *power of a statistical test*).

[14] Alguns testes têm mais poder do que outros quando aplicados ao mesmo conjunto de dados.

[15] Veja, por exemplo: VIEIRA, S. *Introdução à Bioestatística*. 5. ed. Rio de Janeiro: Campus-Elsevier, 2016.

[16] Veja, por exemplo: VIEIRA, S. *Análise de variância*. São Paulo: Atlas. *E-book*.

[17] Veja, por exemplo: VIEIRA, S. *Introdução à Bioestatística*. 5. ed. Rio de Janeiro: Elsevier, 2016.

18 Veja a seção 5.2 do Capítulo 5 deste livro.

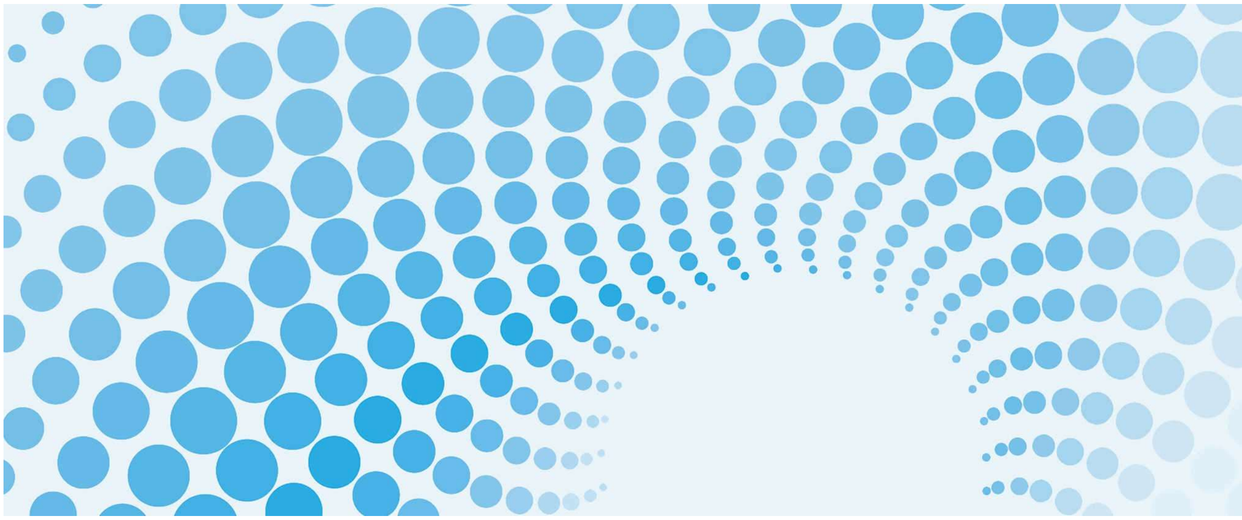
[19] Veja o procedimento para o teste *t* de Student em VIEIRA, S. *Introdução à Bioestatística*. 5. ed. Rio de Janeiro: Elsevier, 2016. Capítulo 13.

[20] Id., ibid.

[21] Veja o Capítulo 5 deste livro.

[22] Veja o procedimento para análise de variância em VIEIRA, S. *Análise de variância*. São Paulo: Atlas, 2006.

[23] Id., ibid.



2.5.1. Construa as hipóteses no caso de pesquisa com o objetivo de: a) verificar se uma nova droga é melhor do que a tradicional; b) verificar se determinada dieta aumenta a longevidade; c) verificar se um produto é cancerígeno; d) verificar se uma vitamina aumenta o desempenho de atletas.

2.5.2. Para comparar a prevalência de alcoolismo entre homens e mulheres residentes do Rio de Janeiro, foram levantados os dados apresentados na Tabela 2.2. Construa as hipóteses. Foi feito o teste estatístico denominado qui-quadrado. O *p*-valor é 0,0004. O que você conclui?

Prevalência de alcoolismo segundo o sexo

Sexo	**Alcoolismo**		**Total**
	Sim	**Não**	
Masculino	29	559	588
Feminino	15	856	871
Total	44	1.415	1.459

Fonte: ALMEIDA, LM, COUTINHO, ESF. Prevalência de consumo de bebidas alcoólicas e de alcoolismo em uma região metropolitana do Brasil. Rev Saúde Pública 27 (1), 1993.

2.5.3. Um pesquisador pretende conduzir um experimento para testar uma hipótese[24]. Se resolver duplicar o tamanho da amostra, qual dos seguintes itens irá aumentar?

a) O poder do teste.

b) O efeito do tratamento.

c) A probabilidade de erro Tipo II.

Escolha a resposta:

1. O poder do teste.
2. O efeito do tratamento.
3. A probabilidade de erro Tipo II.
4. Todos os itens listados.
5. Nenhuma das alternativas listadas.

2.5.4. Para verificar se meninos e meninas da mesma idade têm igual velocidade de leitura e igual capacidade de interpretação de textos, um pesquisador pediu a um grupo de crianças que lessem determinada poesia. Para cada criança, cronometrou o tempo de leitura e atribuiu um conceito para medir a capacidade de interpretação. Você aplicaria um teste paramétrico para analisar qual das variáveis?

2.5.5. Você quer testar a hipótese de que uma moeda é bem balanceada. Quais são as hipóteses em teste? Você resolve jogar a moeda seis vezes e dizer que não é bem balanceada se aparecerem seis caras. Um estatístico calcula que a probabilidade de isso acontecer, quando a moeda é bem balanceada,[25] é

$$\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}=0{,}015625$$

Qual é o *p*-valor?

2.5.6. Dos 100 pacientes submetidos a determinada cirurgia, oito vieram a óbito. Que população você definiria para essa estatística?

2.5.7. Poder do teste significa:[26]

a) Não rejeitar H_0 quando H_0 é verdadeira.

b) Não rejeitar H_0 quando H_0 é falsa.

c) Não rejeitar H_0 independentemente de H_0 ser falsa ou verdadeira.

d) Rejeitar H_0 quando H_0 é verdadeira.

e) Rejeitar H_0 quando H_0 é falsa.

2.5.8. Foi feito um ensaio para comparar o efeito analgésico de duas drogas, A e B, e aplicado um teste estatístico para comparar os resultados. Quais são as hipóteses em teste?

2.5.9. Foi medido o nível de determinado hormônio em 50 mulheres jovens que não estavam em gestação e em 50 mulheres jovens que estavam no primeiro trimestre de gestação. Foram obtidas as médias, 93 e 110, respectivamente, e obtido um *p*-valor para um teste unilateral igual a 0,001, porque havia informações na literatura de que gestantes teriam maior quantidade desse hormônio. Escreva as hipóteses e interprete o *p*-valor.

2.5.10. Escolha uma das alternativas abaixo para interpretar a expressão “resultado significante”:

a) A pesquisa é de boa qualidade.

b) O pesquisador aceitou a hipótese da nulidade com probabilidade de 0,5.

c) A probabilidade de o estudo ser verdadeiro é maior do que 5%.

d) A probabilidade de o pesquisador ter obtido o resultado que obteve por acaso é⌊0,05.

2.5.11. Quais dos seguintes procedimentos reduz o poder de um teste de hipóteses[27]?

a) Aumentar o tamanho da amostra.

b) Aumentar o nível de significância.

c) Aumentar a probabilidade de erro Tipo II.

Escolha a resposta:

a) Aumentar o tamanho da amostra.

b) Aumentar o nível de significância.

c) Aumentar a probabilidade de erro Tipo II.

d) Todos os itens listados.

e) Nenhum dos itens listados.

---

[24] *Power of a Hypothesis Test*. Disponível em: http://stattrek.com/hypothesis-test/power-of-test.aspx?Tutorial. Acesso em 24 de outubro de 2017.

[25] Veja o Capítulo 9 de VIEIRA, S. *Introdução à Bioestatística*. 4. ed. Rio de Janeiro: Campus/Elsevier, 2008.

[26] Prova: Analista do CNMP (Conselho Nacional do Ministério Público), 2015.

[27] *Power of a Hypothesis Test*. Disponível em: http://stattrek.com/hypothesis-test/power-of-test.aspx?Tutorial. Acesso em 24 de outubro de 2017.