4.2. Introdução aos testes de hipóteses científicas

No início, a Estatística era útil principalmente para descrever a situação de um "Estado" (seja um país, uma província, uma cidade, ou qualquer outra unidade política), como a própria palavra "Estatística" indica. As ciências, especialmente as que eram chamadas de "exatas", como a Física e a Química, se desenvolveram rapidamente a partir do século XVI, praticamente por conta própria - na verdade, havia muito pouco na Estatística de então que pudesser útil nestas áreas. A partir do final do século XIX, porém, a situação mudou completamente; hoje, a Estatística não é mais considerada simplesmente um conjunto de técnicas usadas para "moer" números – isto é, para organizá-los em tabelas e calcular médias, etc., e sim, uma parte essencial do método científico:

> Estatísticos não se consideram "moedores de números"; mais frequentemente, eles se consideram como os guardiães do método científico. Rumsey (2003). [[1]]

> Estatística (...) é parte da corrente principal da metodologia da ciência. O gênio de Sir Ronald Fischer foi que ele não se via apenas como um estatístico, mas antes de tudo como um cientista. Jenkins (1979). [[2]]

Sem Estatística, não é possível planejar experimentos e analisar seus resultados para testar hipóteses, nos vários ramos em que hoje se subdividiu a ciência. Para entender por que isso ocorre, teremos que discutir o que é uma *hipótese* científica, e como elas podem ser *testadas*.

Uma hipótese científica é uma tentativa de explicar como alguma coisa acontece – desde como um planeta se movimenta em torno do sol, até como uma bactéria rompe as barreiras do sistema imunológico de uma pessoa. Nem toda explicação, porém, é considerada "científica"; existe uma área de pesquisa, a Filosofica da Ciência, que estuda justamente como a Ciência funciona, e o que torna uma teoria "científica" (e não uma pseudociência, ou uma superstição). Segundo estes filósofos, existem várias características que uma hipótese deve ter para que possa ser considerada científica; contudo, - como a Filosofia, em qualquer de suas áreas, está longe de ser uma ciência exata -, existe muita discordância entre eles. Em geral, porém, todos concordam num ponto: uma teoria científica tem quer ser *testável*; isto é, deve ser possível projetar e executar um experimento que possa *testar* a hipótese, e mostrar quanto de verdade existe nela.

Às vezes, a testabilidade ou não de uma teoria depende da tecnologia disponível num dado momento. Hipóteses não testáveis eram feitas, por exemplo, para explicar o que deveria existir no lado oculto da lua (a lua gira mostrando sempre a mesma face em direção à Terra - o que está do outro lado não pode ser visto por nós). Havia teorias que diziam que lá existiam por exemplo bases de discos voadores pertencentes a civilização extra-terrestres. Os astrônomos, porém, não se importavam por tais teorias – consideravam que não valia a pena perder tempo com teorias que não poderiam nunca ser testadas. Depois que começaram as viagens espaciais, nos anos 1960, a face oculta da lua foi fotografada, e estas teorias passaram a ser testáveis – e o que se verificou foi que o lado oculto da lua é exatamente como o lado visível.

Outras teorias talvez nunca serão testáveis. Por exemplo, existem várias teorias para explicar como surgiram as diversas línguas faladas pelos humanos. Existe a teoria de que a línguagem começou de grunhidos e interjeições, e a partir daí, se organizou até resultar nas complexas gramáticas que governam as línguas modernas (Schopenhauer, 2019). Esta teoria não pode ser testada experimentalmente num laboratório, e os linguistas passaram muito tempo procurando alguma língua "primitiva" que pudesse ser uma evidência a seu favor – alguma língua que tivesse uma gramática feita por regras obviamente rudimentares e incompletas, representando um estágio intermediário na evolução dos grunhidos até as línguas modernas. Esta busca não deu nenhum resultado; o que se descobriu foi que mesmo os povos mais primitivos do ponto de vista tecnológico, como os aborígenes da Austrália ou da África do Sul - que vivem exatamente como seus ancestrais viviam a 40.000 atrás – já possuem línguas tão complexas em suas estruturas quanto qualquer língua moderna. Teorias sobre a origem das línguas, portanto, não podem ser testadas – não existem evidências nem contra nem a favor delas – e portanto são consideradas não-científicas (e geralmente não são discutidas em congressos ou revistas de Linguística.

4.2.1. Como testar uma teoria científica?

As primeiras teorias científicas importantes, surgidas a partir do Renascimento, eram relativamente fáceis de testar. Galileu, por exemplo, criou a teoria de que todos os corpos caem com a mesma velocidade, o que contradizia a filosofia de Aristóteles. Aristóteles, que foi a base de todo o conhecimento medieval, afirmava o que nos parece ser mais intuitivo: os corpos mais pesados caem com maior velocidade. Galileu porém, foi mais longe, e fez algo que Aristóteles nunca tinha feito: realizou um experimento para testar empiricamente sua teoria. Diz a tradição que ele subiu ao topo da torre inclinada de Pisa, e deixou cair de lá duas bolas de bronze, de mesmo diâmetro mas pesos diferentes, e verificou que as duas chegaram ao chão no mesmo instante; isto foi uma evidência a favor de sua teoria, e contra a de Aristóteles.

Outra teoria revolucionária surgida na mesma época foi a de Kepler, que afirmava que os planetas se movem em órbitas elípticas, novamente contradizendo Aristóteles, que afirmava que as órbitas deveriam ser circulares. Testar uma teoria destas já era um pouco mais complicado: não é possível ao astrônomo, sentado num observatório aqui na Terra, observar as trajetórias dos planetas, e ver se são elipses ou não; trajetórias não são diretamente observáveis.

Para qualquer teoria, o que o cientista deve fazer é *deduzir* alguma conseqüência da teoria (o que aconteceria se a teoria fosse verdadeira), e depois verificar empiricamente, por experimentos ou por observação, se esta conseqüência prevista realmente ocorreu. Na teoria de Galileu, a dedução era fácil: se os corpos caem todos com a mesma velocidade, devem chegar ao solo no mesmo instante, se caírem de uma mesma altura. Para a teoria de Kepler, a dedução era bem mais difícil; Kepler deveria ajustar um modelo de elipse às posições passadas de um planeta, e, por extrapolação, prever a sua posição num momento futuro. Kepler conseguiu prever o trânsito de Mercúrio (a passagem deste planeta em frente ao Sol, num dia em 1631). Para isto, precisou de dados bastante acurados sobre a posição de Mercúrio no passado; o teste de sua teoria só foi possível porque estes dados, obtidos com acurácia até então nunca vista, tinham sido recentemente obtidos por Tycho Brahe.

Testes podem fazer que uma teoria seja rejeitada, e substituída por outra (como as de Aristóteles, que foram rejeitadas por Galileu e Kepler; ou a de Ptolomeu, que afirmava que o Sol girava em torno da Terra, foi rejeitada por Copérnico, que afirmava o contrário). Testes podem também fazer com que uma teoria passe a ser considerada caso particular de uma teoria mais geral. (Foi o que aconteceu com as teorias de Galileu e de Kepler, que passaram a ser consideradas casos particulares de uma teoria maior, a lei da Gravitação de Newton.) Quando mais abrangente uma teoria científica, mais útil ela é para a Ciência – e mais difícil de ser testada. A lei de Newton é mais abrangente do que as de Galileu e de Kepler, porque explica não apenas a queda dos corpos perto da Terra ou o movimento dos planetas, mas também outros fenômenos, como a ocorrência das marés, ou a forma esférica dos planetas e estrelas. Contudo, por ser tão abrangente, ela é mais difícil de se testar.

Vejamos como exemplo uma teoria ainda mais complexa que a da Gravitação - a teoria da Relatividade (que o próprio Einstein considerava como uma extensão da teoria de Newton). Esta teoria diz que a massa de um objeto distorce o espaço em seu redor; se isto for verdade, uma das conseqüências é que a trajetória da luz deve parecer se curvar quando os raios luminosos passam perto de um corpo de grande massa. De fato, em 1919, uma equipe de astrônomos conseguiu obter fotos de estrelas que estavam *atrás* do sol (estas fotos foram feitas em Sobral, no Ceará). Estas estrelas deveriam ser invisíveis, se a luz se movesse sempre em linha reta, porque o sol bloquearia os raios luminosos; no entanto, os raios se curvaram em torno do sol, e atingiram a Terra. Esta descoberta se tornou umas primeiras fortes evidências em favor da teoria da Relatividade.

Neste ponto, vale a pena comentar sobre um detalhe: por que a teoria de Newton é geralmente chamada *Lei* da Gravitação, enquanto a de Einstein é chamada mais modestamente de *Teoria* da Relatividade? Qual a diferença entre uma lei e uma teoria, na ciência? Na verdade, não há nenhuma. A "lei" de Newton recebeu este nome, no século XVIII, porque na época ainda se acreditava que a Física tinha a capacidade de descobrir as *leis* que governam o funcionamento do universo, tão exatas e imutáveis quanto os teoremas matemáticos. Hoje em dia, somos mais modestos, e acreditamos que qualquer teoria científica é apenas isto: uma *teoria*, uma tentativa de explicar o funcionamento do universo. As teorias atuais são as melhores explicações de que dispomos no estágio atual da ciência, mas não acreditamos que elas sejam perfeitas e eternas; o trabalho dos cientistas é o de tentar completar os detalhes faltantes de uma teoria, ou corrigir suas imperfeições – ou simplesmente descartá-la, e substituí-la por uma teoria nova, se as evidências empíricas mostram que esta teoria nova tem mais poder explicativo do que a teoria anterior.

4.2.2. Dedução x Inferência

Esta é, em termos muito simplificados, a forma como uma teoria científica é testada: uma conseqüência da teoria é deduzida, e um experimento ou observação é projetado para verificar se esta conseqüência realmente ocorreu. Na prática, existem várias complicações – hipóteses auxiliares que devem ser levadas em conta, variáveis de confundimento (ver seção **2.2.4.3**), erros de medição ou observação, etc. – mas estas complicações não modificam a estrutura básica.

Em termos de Lógica, toda teoria pode ser escrita como uma *implicação*: se A acontece, então B acontece. O processo que nos permite chegar a esta conclusão é chamado de *dedução*. Esquematicamente, o processo pode ser representado como na Fig. **1**.

se acontecer o fenômeno A	→ (dedução) →	vai acontecer o fenômeno B

Figura 1. Exemplo de dedução

A Matemática progride por meio de cadeias de deduções; cada teorema tem que ser deduzido logicamente a partir de definições e axiomas iniciais, ou de teoremas já deduzidos anteriormente (uma dedução matemática é chamada de *demonstração*). A Matemática não faz experimentos ou observações para testar suas hipóteses (que são chamadas de *teoremas*), mas trabalha exclusivamente por meio de deduções.

Para um cientista, porém, a dedução é apenas o primeiro passo. O cientista deduz que, se a hipótese A for verdadeira, deve ocorrer a conseqüência B - por exemplo, se a Teoria da Relatividade for verdadeira, a luz deve ser curvar ao passar perto do Sol. O que mais interessa ao cientista, porém, é o caminho inverso: se B ocorreu, que posso concluir sobre A? (Fig. **2**).

se a hipótese A for verdadeira	→ (dedução) →	vai acontecer a conseqüência B
		↓
		(experimento ou observação)
		↓
a hipótese A é verdadeira (?)	← (inferência) ←	se aconteceu a conseqüência B

Figura 2. Dedução e inferência na Ciência

Se a luz realmente se curvou, que posso concluir sobre a Teoria da Relatividade? Na Lógica, este processo inverso é chamado de *indução* ou de *inferência* (as duas palavras significam a mesma coisa) [[3]]. A dificuldade é que é a dedução é sempre *certa* e definitiva. Se um teorema matemático foi deduzido, ele não poderá jamais ser rejeitado. O teorema de Pitágoras, por exemplo, demonstrado por volta do século VI AC, continua válido até hoje. A *inferência*, porém, nunca tem o mesmo poder; o resultado de um experimento nunca é considerado como *prova* definitiva de que uma teoria esteja certa. O experimento sobre a curvatura da luz não *prova* que a teoria da Relatividade seja verdadeira; ele é apenas, mais modestamente, um *evidência* (embora, neste caso, bastante forte) de que a teoria *pode* ser correta. Este processo pode ser representado esquematicamente como na Fig. **3**.

se a teoria da Relatividade for *verdadeira*	→ (dedução)	os raios luminosos irão se curvar ao passar perto do sol
		↓
		(experimento)
		↓
há uma forte evidência de que a teoria da Relatividade é *verdadeira*	← (indução / inferência)	se os raios luminosos realmente se curvaram ao passar perto do sol

Figura 3. Dedução e indução num experimento

Teorias científicas podem ser *corroboradas* por experimentos, mas nunca realmente *provadas*. Se um experimente foi aparentemente bem sucedido, pois comprovou a ocorrência do fenômeno B, isto poder ter ocorrido porque a hipótese A era realmente verdadeira, mas também pode ter ocorrido em conseqüência de diversos outros fatores, que não têm nada a ver com a hipótese. Um destes fatores é simplesmente o erro de medição ou de observação. Na pesquisa, os cientistas trabalham quase sempre no limite de precisão dos instrumentos de que dispõem, e erros ou interpretações erradas sempre podem acontecer. (Tycho Brahe por exemplo não tinha um telescópio, e fez a olho nu todas as observações de planetas que Kepler usou depois em seus cálculos. Kepler aliás já tinha em telescópio, mas de qualidade tão duvidosa que ele concluiu que o planeta Marte era quadrado.)

Um exemplo famoso destes erros foi a polêmica que surgiu no início de século XX, quando o astrônomo Percival Lowell publicou mapas dos “canais” que teria visto na superfície de Marte, evidentemente artificiais (porque eram retilíneos, o que não ocorre com os rios e lagos naturais), e concluiu daí que o planeta tinha habitantes inteligentes, com tecnologia avançada. Vários astrônomos negaram a existência destes canais, mas alguns apoiaram Lowell por muito tempo; a discussão só foi realmente decidida depois que as sondas *Mariner* enviadas pela NASA fotografaram de perto a superfície de Marte, e mostraram que os canais simplesmente não existem.

Um outro fator que dificulta a inferência, é que o fenômeno B talvez possa ser consequência de outras causas, não relacionadas com a hipótese A. Um exemplo simples: se a escala no braço de um violão é mal feita (os trastes não foram colocados nas posições corretas), as notas soarão desafinadas. Se alguém toca um violão e ele soa desafinado, posso concluir que a escala é mal feita? Não, porque o problema pode estar nas cordas – cordas de baixa qualidade, que não têm uniformidade na densidade e no diâmetro, não podem ser afinadas. Ou pode estar também no músico – a colocação dos dedos na escala, e a pressão feita, também podem alterar um pouco a afinação.

Na verdade, a aceitação ou rejeição de uma teoria é em geral um processo muito longo, que envolve muita discussão. Na Física, que é provavelmente a mais “exata” das ciências, se temos um resultado como este que mostra a curvatura da luz, obtido em experimentos bem planejados e executados (a possibilidade de ter ocorrido algum erro - por exemplo, fotografar as estrelas erradas... - pode ser descartada), e não existem outros fatores que possam explicar o fenômeno B (ninguém encontrou nenhuma outra explicação possível para a curvatura da luz), chegamos o mais próximo que se pode chegar da certeza.

Contudo, qual teria sido a que conclusão, se os raios luminosos *não* tivessem se curvado, e as tais estrelas *não* fossem visíveis da Terra? Neste caso, se temos a certeza de que não houve nenhum erro no experimento, concluímos que há uma evidência muito forte, quase uma certeza, de que a teoria *não* é verdadeira (Fig. **4**).

se a teoria da Relatividade for *verdadeira*	→ (dedução)	os raios luminosos irão se curvar ao passar perto do sol
		↓ (experimento) ↓
a teoria da Relatividade **não** é *verdadeira*	← (indução / inferência)	se os raios luminosos **não** se curvaram ao passarem perto do sol

Figura 4. Dedução e indução num experimento, levando à rejeição de uma hipótese

Existe portanto uma assimetria nas conclusões retiradas a partir de experimentos: se B ocorreu, isto não prova que a teoria A seja verdadeira (embora possa ser uma *evidência* a favor de A); se B não ocorreu, porém, isto praticamente uma prova que a teoria A é falsa.

4.2.3. Teorias sobre populações e amostras

Nesta altura, você deve estar se perguntando: tudo isto pode ser interessante, mas o que tem a ver com Estatística? Veremos agora teorias que se referem a populações a amostras, que são as que mais interessam à Estatística. A hipótese é geralmente uma afirmativa sobre a população. O que o cientista deve fazer é deduzir: o que espero encontrar na amostra, se a hipótese for verdadeira?

Consideremos um exemplo simples: suponha a hipótese determinística de que “todo brasileiro gosta de samba”. Se isto for verdade, e retiro uma amostra de 100 brasileiros, é de se esperar que todos eles gostem de samba. Se, depois de interrogar as 100 pessoas, verifico que todas de fato gostam de samba, posso concluir que a hipótese é verdadeira? Não; talvez *muitos*, ou a *maioria*, ou 99,9% dos brasileiros gostem de samba, mas este resultado não prova que *todos* gostem. Se, por outro lado, encontro *uma* pessoa na amostra que não gosta de samba, isto é suficiente para que eu rejeite a hipótese. Karl Popper, um dos mais influentes filósofos da Ciência do século XX, propôs um exemplo que se tornou clássico: Um pessoa que morava na Europa no século XVIII poderia ter observado centenas de cisnes, e chegado à conclusão (teoria) de que todo cisne é branco. Se esta pessoa, passeando por um parque, encontrasse um cisne branco, isto provaria sua teoria? Não. Se esta pessoa porém fosse para a Austrália (que ainda não tinha sido explorada pelos ingleses), o primeiro cisne que veria seria provavelmente negro – e este único cisne seria capaz de derrubar a teoria. Este exemplo deu origem a expressão “cisne negro”, muito usada nas discussões sobre ciência – um cisne negro é um evento que, se ocorrer, tem o poder de destruir uma teoria.

Popper sugeriu que um cientista deve procurar fazer experimentos que visem não *provar* a teoria que interessa, mas sim rejeitá-la – isto é, deve procurar encontrar um cisne negro. Esta idéia não é aceita por todos os filósofos (como qualquer ramo da Filosofia, a *Filosofia da Ciência* também não é uma ciência, e muito menos uma ciência exata!), mas oferece uma analogia interessante com a Estatística: nos testes estatísticos, em geral é mais fácil, e mais útil, procurar *rejeitar* uma teoria, do que tentar aceitá-la. Por isso, os testes estatísticos têm uma estrutura que a princípio pode confundir os estudantes – se quero testar uma hipótese, devo fazer um experimento que vise testar a hipótese contrária; se consigo rejeitar esta hipótese contrária, terei uma evidência de que a hipótese que me interessa é verdadeira (veremos isto com mais detalhes na Seção 4.3.2)

4.2.4. Inferência na presença de incerteza

Nas ciências onde existe grande incerteza, como a Biologia e a Medicina, a situação é ainda mais complicada (esta é uma das razões porque estas ciências avançaram muito mais lentamente do que a Física e a Química, desde o Renascimento). Nestas ciências, em geral não nos interessam hipóteses como esta de que *todo* cisne é branco, facilmente rejeitada quando um único cisne negro é encontrado. Uma hipótese mais realista, por exemplo, seria a de que *fumar faz mal à saúde*. Depois de dezenas de estudos e de uma discussão que durou mais de 30 anos, é bem aceita hoje a hipótese de que o hábito de fumar é prejudicial à saúde, e diminui a expectativa de vida das pessoas. Se encontramos porém um homem que sempre fumou muito, e chega as 104 anos de idade, saudável e lúcido, podemos concluir que esta hipótese é falsa? (Foi o caso, por exemplo, do arquiteto Oscar Niemeyer). Nestas ciências, não podemos tirar nenhuma conclusão a partir de um único caso, um cisne negro; teremos que usar amostras (conjuntos contendo um grande número de casos), e tirar conclusões *probabilísticas*, não *determinísticas*. Considerando a população de fumantes existentes, tiramos amostras delas, e a partir dos resultados observados concluímos que o hábito de fumar é prejudicial à saúde de uma grande porcentagem dos fumantes (mas não de todos); isto é, tem grande *probabilidade* de causar dano à saúde. (A Física, neste aspecto, é mais simples do que a Medicina: não precisa de criar o conceito de uma *população* de universos diferentes, em alguns dos quais a luz se curva, em outros não ...).

Novamente, o que o cientista deve fazer é deduzir: o que espero encontrar na amostra, se a hipótese for verdadeira? Suponhamos um exemplo muito simples (que será desenvolvido na Seção 4.3.1): queremos testar se uma moeda é equilibrada ou não; isto é, se a pro-babilidade de ela mostrar a *cara* quando lançada é igual a de mostrar a *coroa*. Evidentemente, nada posso concluir depois de fazer um único lançamento. Terei que fazer uma amostra, uma seqüência de lançamentos.

Suponha que eu faça 20 lançamentos, e conte as caras. Que espero encontrar nesta amostra? Qualquer número de caras (de 0 a 20) é teoricamente possível; porém, se a moeda for mesmo equilibrada, é mais provável que eu encontre em torno de 10 caras. Se encontro cerca de 10 caras, posso considerar que isto é uma evidência (embora muito fraca, como veremos depois) a favor da hipótese de que a moeda é equilibrada. Se porém encontro um número muito diferente de 10 – por exemplo, 19 ou 20 caras -, que posso concluir? Precisaremos de um *modelo probabilístico* que nos permita calcular as probabilidades dos diversos valores do número de caras. Este modelo é a *distribuição amostral* – o modelo que nos permite calcular as probabilidades do que pode acontecer na amostra, se a hipótese for verdadeira. Neste exemplo, o modelo será o da distribuição binomial (seção 3.3.5); por meio dele, verificamos que os valores de X (19 ou 20) seriam extremamente improváveis, se a moeda fosse equilibrada ($P \cong 0{,}0000$); é muito mais provável, portanto, que ela seja desequilibrada. (O conceito de "distribuição amostral" é a base de toda a Inferência Estatística; voltaremos a ele nas seções 4.5.1 e 4.5.3).

Este tipo de problema – teste de uma hipótese sobre uma *proporção* ou uma *probabilidade* – é muito comum em diversas áreas. Por exemplo, suponha que o tratamento usual para uma certa doença é bem sucedido em 80% dos pacientes. Um novo tratamento é desenvolvido, e pesquisadores decidem verificar se ele é mais eficiente do que o tratamento usual (isto é, se ele será bem sucedido em *mais* de 80% dos pacientes). Para isto, terão que fazer um teste estatístico, usando uma amostra de pacientes. Um segundo exemplo, em outra área: suponha que uma fábrica produz peças, das quais se sabe que em geral 3% apresentam alguma forma de defeito, e têm que ser descartadas. Uma fábrica concorrente cria um novo processo de fabricação, e afirma que apenas 1% das peças produzidas por este novo processo são defeituosas. Para testar esta afirmação, será preciso fazer um teste estatístico, usado amostras destas peças.

A dificuldade conceitual destes testes estatísticos é que eles nunca chegam a conclusões *exatas e definitivas*, mas sempre a respostas *mais ou menos prováveis*. Comparemos os testes de uma teoria numa ciência "exata" com o de uma teoria numa ciência onde haja grande incerteza. Para a ciência exata, tomemos como exemplo o teste da teoria da Relatividade mencionado acima. Neste teste, partindo da hipótese de que a teoria seja verdadeira, *deduzimos* que os raios luminosos devem se curvar ao passarem perto do sol. Se fizermos um experimento e descobrirmos que os raios não se curvaram, *inferimos* que a teoria é falsa.

Numa ciência "incerta", como a Medicina (William Osler, um médico canadense, definiu a Medicina como "a ciência da incerteza e a arte da probabilidade"), a situação é bem diferente. Suponhamos o exemplo mencionado acima, um teste que visa verificar a hipótese de que um novo tratamento é bem sucedido em *mais* de 80% dos pacientes. Se esta hipótese for verdadeira, quando fizermos um teste com uma amostra de 200 pacientes iremos esperar que o tratamento seja bem sucedido em *mais* de 160 pacientes. Se o tratamento não tiver sucesso em mais que 160 pacientes, chegaremos à conclusão de que a hipótese é falsa; isto é, de que o novo tratamento não é mais eficiente do que o tratamento usual. Esquematicamente, podemos representar este processo como na Fig. **5.**

se a hipótese de que o tratamento obtém sucesso em mais de 80% dos pacientes for verdadeira	→ *dedução*	numa amostra de 200 pacientes, o tratamento deve obter sucesso em *mais de* 160 pacientes
		↓ (experimento) ↓
a hipótese *provavelmente* é falsa; isto é, o novo tratamento *provavelmente* **não** é mais eficiente que o tratamento usual	← *indução* ou *inferência*	se numa amostra de 200 pacientes o tratamento **não** obteve sucesso em *mais de* 160 pacientes

Figura 5. Dedução e indução num experimento, com base em probabilidades

Note que há duas incertezas neste processo. Em primeiro lugar, a expressão "mais de 160 sucessos" é vaga. Se encontro na amostra 164 pacientes que tiveram sucessos no tratamento, posso concluir que o novo tratamento é mais eficiente do que o tratamento usual? Provavelmente não; o tratamento antigo também teria uma probabilidade razoável de conseguir um resultado igual ou melhor do que este (p=0,215; veremos como calcular isto na Seção 4.5.2). Um resultado destes não traz evidência suficiente para concluirmos que haja alguma diferença entre os dois tratamento (ele é chamado de *não-significativo*). Por outro lado, se encontro na amostra 190 pacientes que obtiveram sucesso no tratamento, a conclusão será bem diferente; a probabilidade de o tratamento antigo conseguir um resultado igual ou melhor do que este é praticamente nula (p=0,000). Irei portanto concluir que é *provável* que tratamento novo seja mais eficiente do que o antigo; este resultado será chamado de *significativo*.

A segunda incerteza está nas conclusões: concluímos que o novo tratamento é *provavelmente* mais eficiente do que o anterior. Quão *provável* é esta conclusão? Qual é a probabilidade de esta conclusão estar *errada*? Veremos a seguir as técnicas de inferência estatística fornecem métodos para calcular a margem de erro nas estimativas feitas; além disso, permitem que os pesquisadores, antes de tomarem a decisão entre aceitar ou não uma hipótese (como a de que o tratamento novo é melhor do que o antigo), calculem qual o *risco* (a probabilidade) de estarem cometendo um erro. Este tema será abordado na Seção 4.3.2.2.

Notas

[1] *Statisticians don't think of themselves as number crunchers; more often, they think of themselves as the keepers of the scientific method.*

[2] *Statistics (...) is part of the mainstream of the methodology of science. The genius of the late Sir Ronald Fisher was that he saw himself not only as a statistician but first and foremost as a scientist.*

[3] A *indução matemática*, que é ensinada nas escolas, não tem na verdade nada a ver com o que é chamado de *indução* em Lógica; é na verdade uma forma de raciocínio baseado numa cadeia de *deduções*.

Referências

Jenkins, G. M. (1979). *Practical experiences with modelling and forecasting time series.* St. Helier, Jersey, Channel Islands: Gwilym Jenkins & Partners.

Rumsey, Deborah (2003). *Statistics for Dummies*. For Dummies edit.

Schopenhauer, A. (2019). *A arte de escrever*. Porto Alegre: L&PM. p.145.