

UMA INICIAÇÃO AOS TESTES ESTATÍSTICOS PARA DADOS BIOLÓGICOS

INCLUI SCRIPTS DOS TESTES EM R



RONEI BALDISSERA

UMA INICIAÇÃO AOS TESTES ESTATÍSTICOS PARA DADOS BIOLÓGICOS

INCLUI SCRIPTS DOS TESTES EM R

RONEI BALDISSERA



À Rosana, com amor...

“Eles passarão, eu [a gente] passarinho.”
Mário Quintana

SUMÁRIO

PREFÁCIO	05
1 DADOS BIOLÓGICOS E SUA ANÁLISE	08
2 VARIÁVEIS	24
3 DISTRIBUIÇÕES DE PROBABILIDADE	32
4 TESTE DE HIPÓTESE	39
5 ESTATÍSTICA DESCRITIVA DE VARIÁVEIS QUANTITATIVAS	51
6 TESTES ESTATÍSTICOS	61
7 MISCELÂNEA	110
8 ANÁLISES NO R	121
REFERÊNCIAS	151
GLOSSÁRIO	156
SOBRE O AUTOR	159

PREFÁCIO

O principal objetivo deste livro é ser uma ferramenta para auxiliar estudantes de cursos superiores, principalmente da área das Ciências Biológicas. É um pressuposto desejável que o estudante tenha sido apresentado aos conceitos básicos envolvidos em uma pesquisa científica, seja em uma disciplina regular, tal como metodologia científica ou introdução à estatística, ou em um projeto de iniciação científica. O texto pode ser utilizado como referência complementar nos estudos da análise de dados biológicos. O objetivo é que os estudantes saibam como escolher um teste estatístico para um experimento específico, como aplicá-lo e como interpretar os resultados. As bases estatísticas apresentadas também podem ser utilizadas por pessoas que cursam outras disciplinas. Porém, não serão abordadas profundamente as bases matemáticas dos testes. Para muitos estudantes, a estatística é uma ferramenta útil e conhecer as bases matemáticas detalhadas de um teste estatístico não fará com que os resultados de análises sejam melhores ou piores.

Por outro lado, é importante ressaltar que “testar dados” pressupõe que eles já estejam disponíveis. Entretanto, não é objetivo ensinar técnicas de coleta e de desenho amostral. Cada disciplina das Ciências Biológicas possui seus próprios métodos de coleta. Desenho amostral é um conceito que se refere a um conjunto de procedimentos que descreve como a coleta dos dados é realizada e se baseia no tipo de pesquisa: exploratória, descritiva ou explicativa. O desenho amostral é tão ou mais importante do que a análise, pois é a fonte da informação que será analisada. Se a fonte for deficiente, as análises refletirão essa deficiência e as evidências científicas poderão ter pouca confiança.

O livro está organizado em oito partes, de forma que o estudante possa consultá-lo rapidamente para uma análise específica. A primeira parte aborda a natureza dos dados biológicos e os tipos de estudos biológicos. Também são abordados os conceitos de universo amostral e amostra. A definição de estatística é feita nesta parte, bem como as diferentes abordagens estatísticas que são utilizadas atualmente e a inferência estatística. Ao final desta parte, se mostra um exemplo prático dos passos de uma pesquisa. A segunda parte apresenta os

tipos de variáveis, suas classificações e a maneira básica de lidar com variáveis de confusão. A terceira parte apresenta as distribuições de probabilidade das variáveis aleatórias, com ênfase na distribuição normal. O teste de hipótese é apresentado na quarta parte. Neste livro, esse teste é a base na qual se assenta a tomada de decisão a partir do resultado do teste estatístico. Atualmente, há muita controvérsia sobre como esse teste é utilizado na ciência. Por isso, também se abordam, na quarta parte, as armadilhas que devem ser evitadas quando se utiliza essa abordagem. A parte cinco apresenta os aspectos descritivos da estatística para variáveis quantitativas, que é o ponto de partida exploratório da análise de dados. Também são apresentadas as estimativas das quais se ouve falar no dia a dia, tal como a média, o desvio-padrão e o intervalo de confiança. Os testes estatísticos para variáveis nominais e quantitativas são apresentados e discutidos na parte seis, em que se abordam os testes mais conhecidos e utilizados em pesquisas biológicas, especialmente os testes chamados univariados. Esses testes formam um conjunto chamado de modelos lineares gerais, no qual se inserem a famosa ANOVA e a regressão linear. Na sétima parte do livro, a miscelânea, abordam-se tópicos diversos, os quais, apesar de estarem à parte, também são importantes para a prática da análise de dados, por exemplo, a lógica de testes de permutação. Esses testes podem substituir os testes com variáveis quantitativas se os dados não seguirem determinadas regras que os deixam inadequados para análise com testes clássicos. Nesta parte, também se encontram os testes chamados não paramétricos mais conhecidos. Na oitava parte, faz-se uma introdução ao uso do pacote estatístico R e apresentam-se os tutoriais para realização dos testes apresentados no livro a partir de dados reais. O estudante praticará os testes e poderá utilizar seus próprios dados para realizar todas as análises. Finalmente, a última parte apresenta um glossário de termos utilizados no livro para consulta rápida.

Para cada teste estatístico demonstrado, a lógica utilizada foi apresentar o teste, para que serve, como se faz e, depois, um exemplo. A maioria dos exemplos foi retirada de estudos reais. Também são apresentados os passos lógicos para cada teste e uma forma de apresentar os resultados. Ressalta-se que o próprio título do livro já indica que o texto é uma iniciação. A expectativa é que os estudantes que o lerem tenham um entendimento mínimo sobre a análise de dados e que isso os incentive a aprofundarem seus estudos.

A maioria dos testes estatísticos aqui apresentados pode ser realizada em planilhas do Excel, mas há vários outros programas estatísticos que também podem ser utilizados. A maioria deles exige, no entanto, algum conhecimento básico sobre o manuseio de planilhas (ou matrizes). Programas estatísticos estão disponíveis na internet, tanto de forma gratuita como pagos. Cada uma das análises deste livro pode ser realizada na maioria dos programas gratuitos ou comerciais disponíveis.

Boa leitura e boas pesquisas!

1 DADOS BIOLÓGICOS E SUA ANÁLISE

A base da metodologia para a análise de dados biológicos deriva de suas relações com o meio ambiente. Nas ciências físicas, por exemplo, há geralmente relações de causa-efeito entre as condições naturais ou experimentais e os resultados das observações ou experimentos. Isso quer dizer que, dado um certo conjunto de condições iniciais, os resultados futuros podem ser previstos (CAIN, 2014). Entretanto, esse tipo de relação determinística é raro nas ciências biológicas.

Geralmente, a partir de um mesmo conjunto de dados, os dados biológicos amostrados podem gerar resultados diferentes devido ao grande número de fatores atuando, muitos dos quais o observador não tem conhecimento. Por exemplo, a variabilidade genética inata do material biológico é uma fonte importante de variação para qualquer estudo biológico. Porém, nem sempre se leva em conta essa variação nos estudos; somente seus efeitos.

Por outro lado, quando observações biológicas são repetidas muitas vezes sob condições similares, as frequências relativas dos resultados tendem a indicar determinadas probabilidades. Ou seja, não há um resultado “verdadeiro” que sempre se repetirá, mas vários resultados mais ou menos prováveis. Dessa forma, quando se diz que a probabilidade de um evento fruto de uma observação (ou experimento) é igual a P , o real significado é que, em uma longa série de repetições dessa observação, é praticamente certo que a frequência do evento será aproximadamente igual a P . Essa abordagem analítica está baseada na teoria probabilística de frequência de ocorrência de um evento (LEGENDRE; LEGENDRE, 2012), também conhecida como abordagem frequentista.

Além da relação determinística e probabilística entre as condições naturais e o resultado de uma observação, um determinado resultado pode depender das estratégias dos organismos e de seu meio ambiente. Esse tipo de relação se denomina estratégica e se analisa com a teoria dos jogos (ROSS, 2014). Outra possível relação é que haja muitos possíveis resultados com previsibilidade em curto tempo e imprevisibilidade em tempos maiores. Essa relação é chamada de caótica e se estuda com métodos da teoria do caos (OESTREICHER, 2007). Por fim, podemos ter total incerteza dos resultados, o que não pode ser analisado objetivamente.

A qualidade dos resultados e a sua interpretação nos estudos biológicos dependem da compatibilidade entre os dados coletados e os métodos de análise. Mas o que são dados? Um dado é uma informação que descreve o mundo natural, no caso, o mundo natural biológico. Então, um dado é um descritor. Ele traz uma informação a respeito do sistema de estudo.

1.1 TIPOS DE ESTUDOS BIOLÓGICOS

Os estudos biológicos podem ser realizados de formas diferentes em lugares diferentes. Um estudo é a “aplicação da inteligência para adquirir ou ampliar conhecimentos a respeito de algo” (ESTUDO, 2015). Então, estamos estudando continuamente durante nossas vidas. A diferença do estudo descompromissado da vida diária e do estudo científico é a parte da “aplicação da inteligência”. Na ciência, a aplicação da inteligência é feita a partir de um conjunto de procedimentos, ao qual chamamos de pesquisa, que é uma “série de atividades dedicadas a novas descobertas, abrangendo todas as áreas de conhecimento” (PESQUISA, 2015). Esse é um processo sistemático e envolve a coleta, análise e interpretação da informação para entender um fenômeno.

Segundo Leedy e Ormrod (2020), uma pesquisa pode ser dividida nos seguintes tópicos:

1. O cientista começa com um problema – uma questão não respondida – relacionado com um tópico de seu interesse;
2. Ele articula, clara e especificamente, o objetivo da pesquisa;
3. Geralmente, ele divide o problema principal em problemas acessórios que são mais facilmente abordados;
4. O cientista identifica pressupostos gerais – e, possivelmente, também hipóteses específicas – que embasam o esforço de pesquisa;
5. Ele desenvolve um plano específico para abordar o problema e seus problemas acessórios;
6. O cientista coleta, organiza e analisa os dados relacionados ao problema principal e aos problemas acessórios;
7. Por fim, ele interpreta o significado dos dados em relação ao problema principal e aos problemas acessórios.

Neste livro, serão abordados, superficialmente, os itens 5 e 7, e, com mais profundidade, a parte de organização e análise dos dados do item 6. Quanto à parte de coleta do item 6, que tipos de procedimentos podem ser utilizados? A resposta é que depende do problema e dos objetivos da pesquisa. Genericamente, de acordo com Sue; Ritter (2007), os tipos de pesquisa podem ser divididos em:

1. Pesquisa exploratória – pode ser utilizada para formular problemas, esclarecer conceitos e montar hipóteses. Nesse tipo de pesquisa, não é necessário examinar uma amostra aleatória de uma população ou universo amostral. Esse tipo de pesquisa, em geral, objetiva criar hipóteses ao invés de testá-las;
2. Pesquisa descritiva – possui mais diretrizes. Descreve indivíduos, lugares e situações. Em geral, possui uma ou mais questões norteadoras, mas sem possuir hipóteses estruturadas. Este tipo de pesquisa, geralmente, objetiva descrever características de uma população ou universo amostral com base em coletas de amostras. Por isso, com frequência, se utiliza uma técnica de amostragem baseada em probabilidade, tal como a amostragem aleatória simples. Os dados desse tipo de pesquisa podem ser qualitativos ou quantitativos, e os resultados são apresentados em forma de distribuições de frequência ou estatísticas descritivas, tal como a média aritmética;
3. Pesquisa explicativa – procura explicar por que um fenômeno ocorre e prever ocorrências futuras dele. Direciona-se por hipóteses que especificam a natureza e a direção das relações entre as variáveis sendo estudadas. Amostragens baseadas em probabilidades são utilizadas, pois um objetivo é generalizar os resultados para a população ou universo amostral do qual a amostra foi extraída. A informação é quantitativa e, quase sempre, um teste estatístico será utilizado para validar as relações observadas entre as variáveis. Neste livro, serão abordados testes estatísticos que podem ser aplicados às pesquisas descritivas e explicativas.

Seja qual for o tipo de pesquisa escolhida, sempre haverá um objeto de estudo e um ou mais descritores. Descritores são as informações usadas para descrever ou comparar os objetos de estudo. Os objetos que biólogos usam são paisagens, áreas, ecossistemas, parcelas, populações, indivíduos, tecidos, genes, riachos, culturas, talhões, dentre outros. Os descritores usados para descrever

ou qualificar os objetos são as informações físicas, químicas, ou biológicas desses objetos. Descritores são mais conhecidos como variáveis. A Tabela 1 mostra uma classificação de descritores.

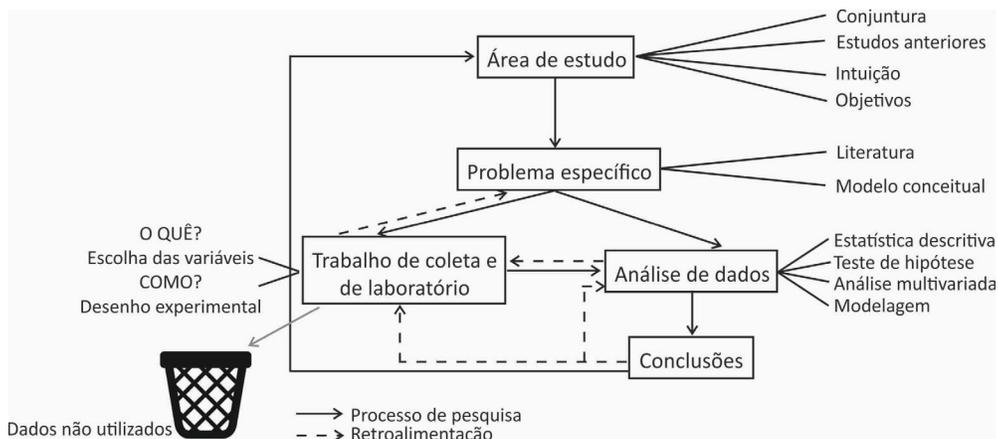
Tabela 1 – Os diferentes tipos matemáticos de descritores

Tipos de descritores	Exemplo ambiental
Binário (dois estados, presença-ausência) Estados múltiplos	Espécie presente ou ausente
Não ordenados (qualitativo, nominal, atributo) Ordenados	Tipos florestais, tratamentos
Semiquantitativos (ordinal)	Importância ou escores de abundância
Quantitativos (métrico, medida)	
Descontínuos (discreto)	Número de filhotes
Contínuos (métrico)	Peso

Fonte: Legendre e Legendre (2012).

É importante levar em consideração os requisitos das técnicas de análise de dados quando se planeja uma pesquisa, pois é inútil coletar dados que são inapropriados para as análises desejadas e que devem ser pensadas a partir dos objetivos do estudo. Coletas de dados mal planejadas podem gerar uma grande quantidade de dados inúteis. A Figura 1 apresenta um fluxograma do processo de pesquisa. As setas cheias representam o processo de pesquisa, enquanto as setas tracejadas caracterizam o processo de retroalimentação. A retroalimentação mostra o progresso científico a partir de um estudo, que gera resultados que serão utilizados para realizar novas coletas dentro do mesmo estudo. Já a seta cheia “novas hipóteses” representa a contribuição do estudo para a literatura científica, criando informações para embasar novos estudos. Sempre que dados gerados não são utilizados, há perda de informação e, portanto, de energia, representada pela lixeira no canto inferior esquerdo da figura.

Figura 1 – Fluxograma do processo de pesquisa científica

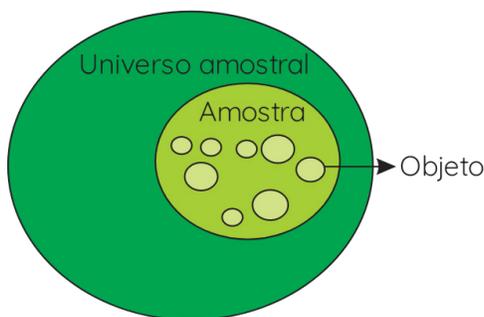


Fonte: Adaptado de Legendre e Legendre (2012).

1.2 UNIVERSO AMOSTRAL E AMOSTRA

Os objetos e descritores devem ser definidos na etapa de desenho amostral da pesquisa. Um número de objetos selecionado é chamado de amostra, a qual é tomada de um conjunto maior, denominado universo amostral ou população estatística. Em cada objeto, será realizada uma medida de um (ou mais) descritor. Quem define o universo amostral é o pesquisador. Por exemplo, uma população de uma espécie de ave dentro de um fragmento de floresta de 10 hectares é o universo amostral, um conjunto de indivíduos dessa população é a amostra e cada indivíduo é o objeto. Na Figura 2, essa característica está representada.

Figura 2 – Representação de universo amostral, amostra e objeto



Fonte: Elaborada pelo autor (2020).

Imaginemos que um pesquisador quer saber o comprimento médio do corpo dessa espécie de ave nesse fragmento em um determinado tempo, digamos uma semana. Esse número é o comprimento médio “real” naquela semana, ou, como se diz na análise de dados, o parâmetro. Difícil de atingir esse objetivo, pois o pesquisador teria de medir “todas” as aves durante esse tempo. Cada parâmetro possui uma estatística correspondente que pode ser medida, com alguma incerteza, a partir da escolha de uma amostra, ou um subconjunto do universo amostral.

O universo amostral está associado à escala do estudo. Se outro pesquisador estiver interessado em amostrar indivíduos da mesma espécie de ave em uma paisagem com 500 fragmentos florestais, o universo amostral aumenta. Nessa lógica, os objetos são as unidades amostrais da pesquisa, que são definidas como a menor unidade em que serão realizadas as medidas dos descritores. Nesse caso, unidade amostral é sinônimo de objeto. Alguns objetos são perfeitamente discriminados, enquanto outros são menos óbvios. Um indivíduo vegetal ou animal, ou um tipo de comportamento, são exemplos de objetos claramente distintos. Um volume de água, ou de solo, ou uma área de floresta são objetos mais arbitrários, pois cabe ao pesquisador determinar quanto de volume ou área consistirá em uma unidade amostral.

1.3 A ESTATÍSTICA E SUAS ABORDAGENS

A estatística é um ramo da matemática utilizado para resumir, analisar e interpretar o que é observado, ou seja, para dar sentido às nossas observações (REID, 2013). A estatística pode ser dividida, basicamente, em estatística

descritiva e estatística inferencial (MENDENHALL; BEAVER; BEAVER, 2020). A primeira é utilizada para descrever um conjunto de dados de modo que não seja necessário ficar olhando para uma planilha cheia de números na tela do computador tentando encontrar algum sentido oculto. A estatística descritiva tipicamente é apresentada na forma de gráficos, tabelas ou como números únicos que resumem um conjunto de dados, tais como a média e o desvio-padrão (BUSSAB; MORETTIN, 2010). Já a estatística inferencial envolve métodos que permitem aos pesquisadores inferir ou generalizar observações tomadas em amostras de um universo amostral (LEEDY; ORMROD, 2020). Nesse sentido, não é necessário coletar as informações de todo o universo amostral, mas de apenas uma parte dele – a amostra, exatamente como foi visto no exemplo das aves. A partir desse subconjunto de dados, o objetivo é fazer inferências ou previsões sobre o funcionamento do sistema de estudo como um todo. Na estatística inferencial, aparece o conceito de probabilidade, pois para realizar generalizações e montar modelos de funcionamento de um sistema de estudo a partir da amostragem, é necessário se utilizar de medidas de incerteza.

O cálculo de probabilidade atribui números entre 0 e 1 para eventos incertos visando a representar a probabilidade com que acontecerão (HAND, 2008). Quando uma probabilidade é 1, é certo que o evento ocorrerá, enquanto uma probabilidade de 0 significa que o evento é impossível. Para um evento que “pode” ocorrer, mas que não é certo ou impossível, um número entre 0 e 1 representa a probabilidade dele ocorrer. Uma visão de probabilidade se baseia na proporção de vezes que um evento aconteceria se as circunstâncias fossem repetidas infinitas vezes (DIENES, 2011). Um exemplo é o jogo da moeda. Se assumirmos que uma moeda é justa (não viciada), a proporção de vezes que o lado “cara” cairá se aproximará de um valor específico se jogarmos o jogo infinitas vezes – no caso $\frac{1}{2}$ ou 50%. Esse tipo de interpretação da probabilidade é chamado de frequentista ou clássico (PERKINS; WANG, 2004). Essa abordagem deu origem à técnica utilizada no que é conhecido como teste da hipótese nula. Resumindo, essa abordagem estabelece um limite crítico, baseado em probabilidade, para aceitar ou rejeitar uma inferência sobre o funcionamento do sistema de estudo como verdadeira *antes* de realizar o experimento. Se o resultado é menor do que esse limite, o pesquisador rejeita uma hipótese de referência por ele determinada. É uma abordagem baseada em simplesmente aceitar/rejeitar um resultado com a crença de que a repetição exaustiva do experimento fornecerá um resultado com a mesma probabilidade ou muito próximo a ela. Se a probabilidade é muito pequena, o resultado é significativo, pois se desvia da aleatoriedade. Esse tipo de abordagem é a mais utilizada pelos biólogos.

A abordagem Bayesiana vem ganhando espaço entre os pesquisadores. Para a abordagem Bayesiana, assume-se que há graus de plausibilidade para uma hipótese ou teoria (DIENES, 2011); ou seja, não há um limite determinado anteriormente ao teste. Essa plausibilidade é atualizada depois da coleta e análise dos dados em um experimento. O pesquisador reconhece que o resultado pode ter valores diferentes e usa as distribuições de probabilidades para expressar sua crença sobre cada valor. O foco muda de saber se uma proposição é “verdadeira” para o grau de confiança que se pode ter em uma proposição que não se sabe se é verdadeira ou falsa (PAULINO *et al.*, 2018).

Outra forma de abordar a estatística em um experimento é por meio dos intervalos de confiança. Essa abordagem visa privilegiar uma estimativa de uma amplitude de valores que são mais prováveis de ocorrer. Intervalos de confiança fornecem uma amplitude de valores possíveis e plausíveis para uma determinada estatística que resume o universo amostral, bem como a probabilidade em que essa amplitude cobre o valor real desse universo amostral (DU PREL *et al.*, 2009).

De qualquer forma, valores de testes estatísticos não devem constituir a única fonte de inferência ou conclusão sobre associações ou efeitos. A razão é que a classificação arbitrária de resultados em “significativos” e “não significativos” é desnecessária e danosa para a interpretação de dados.

1.4 INFERÊNCIA ESTATÍSTICA

A inferência estatística pode ser descrita como o processo de acessar a força da evidência sobre se um conjunto de dados é ou não consistente com um determinado mecanismo hipotético que poderia ter produzido as observações realizadas (MAKAR; RUBIN, 2018). Ou seja, é um método para tomar decisões sobre os parâmetros de um universo amostral, baseando-se em uma estatística calculada da amostra desse universo. Na inferência estatística, estamos interessados em estimar a incerteza associada com os dados coletados. Isso permite fornecer uma amplitude provável de valores que devem englobar o valor “real” da variável no universo amostral. O teste de hipótese e os intervalos de confiança são as aplicações da inferência estatística.

Podemos dividir a inferência em dois tipos: da amostra para o universo amostral e do experimento para a causa (MAKAR; RUBIN, 2018). No primeiro tipo, uma amostra de um descritor é retirada aleatoriamente de um universo amostral e as técnicas estatísticas são usadas para descobrir o que pode ser dito

sobre a variável “verdadeira” do universo amostral. No segundo tipo, o efeito de um tratamento ou intervenção particular é acessado. Nesse tipo, as unidades amostrais não são retiradas aleatoriamente do universo amostral, mas são atribuídas aos tratamentos de forma aleatória.

Estudaremos um exemplo simples de como uma inferência estatística pode ser obtida. Ao mesmo tempo, teremos uma noção dos passos envolvidos na análise de dados, uma lógica que será reproduzida ao longo do livro. Esse exemplo está descrito em Cobb (2007). A Tabela 2 mostra o tempo de recuperação pós-operatório, em dias, para sete pacientes que foram divididos aleatoriamente em dois grupos: um grupo controle, com três pacientes, que recebeu cuidados padrões e, um grupo tratamento, com quatro pacientes, que recebeu um novo método de cuidado pós-operatório.

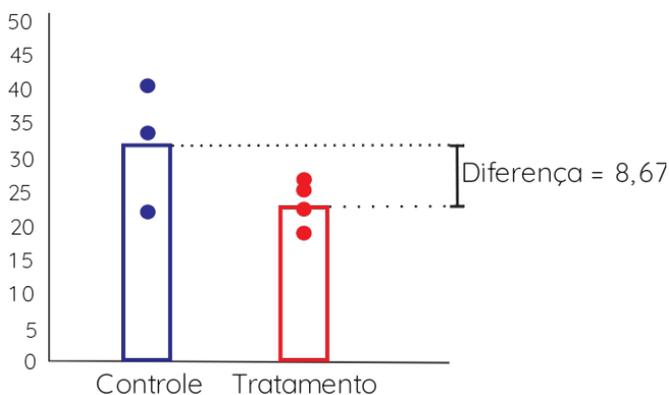
Tabela 2 – Tempo, em dias, de recuperação do pós-operatório em dois grupos de pacientes

	Tempo (dias)				Média de dias	Desvio-padrão
Controle	22	33	40		31,67	3
Tratamento	19	22	25	26	23	9,1

Fonte: Cobb (2007).

A Figura 3 mostra o gráfico com a distribuição dos valores nos dois grupos e o valor da diferença encontrada entre as duas médias observadas.

Figura 3 – Gráfico de barras com a distribuição dos valores e das médias dos dias de pós-operatório



Fonte: Adaptado de Cobb (2007).

Cada método estatístico é derivado de um modelo. Então, o modelo é uma representação da realidade. Nesse caso, o modelo constitui uma única amostra convenientemente coletada, dividida aleatoriamente em dois grupos. Vamos supor que o objetivo seja “saber se o novo método diminui os dias de pós-operatório.” Bem, olhando as médias, é bem óbvio que sim. Entretanto, como esses dados foram retirados de uma amostra (bem pequena) de um universo amostral maior, precisamos aprofundar a pergunta para “A diferença de 8,67 entre as médias é grande o suficiente para ter ocorrido de forma aleatória?” Nesse caso, queremos fazer uma inferência sobre o efeito do tipo de método pós-operatório sobre a quantidade de dias de recuperação dos pacientes.

Um ponto de partida sobre o comportamento destes dados poderia ser que “o tratamento não possui efeito digno de nota”. Podemos dar o nome para essa afirmação de hipótese nula. Neste caso, a palavra nula quer dizer que os sete números e a forma que eles se distribuíram nos dois grupos não tem nada a ver com os efeitos do controle ou o tratamento, pois estamos partindo do princípio de que não há qualquer efeito. A única coisa que determina quais valores foram designados para o grupo tratamento, por exemplo, é a escolha aleatória das pessoas para entrar nesse grupo. Então, a partir disso, podemos realizar um processo chamado permutação, que corresponde ao ato de permutar, ou seja, trocar de lugar, substituir. Isso quer dizer que pegamos todos os sete valores, escrevemos em cartões, os misturamos e retiramos três deles para representar o controle e os outros quatro, representarão o tratamento. Veja que, partindo do princípio de que tudo é igual (nulo), ou não há qualquer efeito do tratamento, podemos misturar todos os valores, sem problemas. Para cada permuta, calculamos os novos valores das médias e a diferença entre elas e anotamos. Ao final de muitas permutas, teremos uma lista com os valores das diferenças para cada uma. Verificamos qual a proporção de vezes em que a diferença entre as médias foi igual ou maior do que 8,67. Se a proporção for pequena, isso quer dizer que a diferença medida no experimento é rara.

Para os dados desse experimento, as 5040 permutas possíveis foram realizadas e houve 756 valores da diferença entre as médias que foram iguais ou superiores a diferença de 8,67 calculada. A proporção então é $756/5040 = 0,15$. Esse é o valor da probabilidade de encontrar um valor da diferença entre as médias igual ou superior ao observado. Ou seja, em torno de 15% das permutações a diferença entre as médias foi igual ou maior do que a observada. Bem, 15% é pouco? É muito? Que decisão tomaremos? Aceitar que essa proporção não é rara o suficiente para inferir que o tratamento mudou a média de dias de

recuperação e, portanto, assumir que a diferença observada foi por pura aleatoriedade? Ou aceitaremos que houve um efeito do tipo de pós-operatório? Neste livro, lidaremos com este tipo de pergunta e aprenderemos como utilizar testes estatísticos que nos auxiliam a decidir que resposta aceitar. Claro, sempre podemos realizar esse processo de permutação por nós mesmos. Os aplicativos de testes estatísticos podem ser relativamente fáceis de programar para reproduzir o processo. Alternativamente, podemos usar aplicativos que já estão pré-configurados para realizar um determinado teste assim que apertamos um botão. Mais importante do que a ferramenta é o entendimento da pessoa que a usa.

1.5 EXEMPLO PRÁTICO DOS PASSOS DE UMA PESQUISA

Aqui, será apresentada de forma sistemática uma maneira de analisar um estudo biológico do tipo pesquisa explicativa. Alguns passos básicos são descritos. Não necessariamente o 3º passo deve vir antes do 4º passo. Uma vez que haja familiaridade com o processo de pesquisa, muitas decisões são tomadas praticamente em conjunto. O exemplo que acompanha os passos é baseado em um estudo biológico na área da ecologia (BALDISSERA *et al.*, 2010). Os autores contaram o número de espécies de plantas em 12 comunidades biológicas em dois campos com diferentes intensidades de manejo: com manejo por gado e sem manejo.

PASSO 1

Especificar a questão ou problema de estudo, o universo amostral e a amostra com os objetos de estudo. A melhor maneira de fazer isso é ler artigos científicos sobre o sistema de estudo e sobre desenho amostral a partir de uma pesquisa bibliográfica para adquirir conhecimento sobre o sistema de estudo. Assim, você terá uma boa ideia de como seus pares determinam essas partes. Uma vez definidos os objetos de estudo, a escolha das variáveis é facilitada.

Com base na breve descrição do estudo, uma questão pode ser “O manejo por gado pode influenciar o número de espécies de plantas campestres?” A questão de estudo é usualmente algo sobre um processo biológico, geralmente na forma “A alteração em X causará alteração em Y?” Pode-se querer saber se uma droga muda a pressão sanguínea; se o pH do solo afeta o crescimento de mudas de uma espécie de planta; ou se uma proteína afeta o transporte de membrana nas células.

PASSO 2

Colocar a questão na forma de uma hipótese e uma previsão de estudo. Quando se tem familiaridade com o assunto da pesquisa, é relativamente fácil determinar que tipo de resposta se espera do sistema de estudo. A diferença entre hipótese e previsão é sutil. A hipótese tem a ver com uma expectativa teórica e a previsão, com uma expectativa operacional ou quantitativa do que se espera como resposta do sistema de estudo.

A hipótese do estudo de exemplo pode ser “O manejo altera substancialmente as condições e recursos ambientais e, portanto, o número de espécies das comunidades vegetais campestres.” No caso, condições e recursos são conceitos ecológicos bem abrangentes. A previsão do estudo declara objetivamente em qual área de campo (com ou sem manejo) se espera encontrar mais espécies. Essa previsão deve estar embasada em informações já coletadas sobre o sistema de estudo ou na teoria. Nesse experimento, a previsão é de que há mais espécies de plantas no campo manejado e essa previsão se baseia na hipótese ecológica de que distúrbios intermediários causam aumento na diversidade de espécies (CONNELL; SLATYER, 1977).

PASSO 3

Colocar a questão na forma de uma hipótese estatística nula e uma hipótese estatística alternativa.

A hipótese nula estatística pode ser “A área com manejo possui o mesmo número médio de espécies que a área sem manejo.” Então, a hipótese alternativa seria “A área com manejo possui um número médio de espécies diferentes da área sem manejo.” No caso da hipótese de estudo, há um processo biológico por trás (algo dentro de condições e recursos). No caso das hipóteses estatísticas, tudo se resume aos resultados numéricos (nesse caso, as médias). Nesse experimento, o número de espécies é estatisticamente diferente ou igual. O teste de hipótese é o principal tema deste livro, e esse teste deve dar uma resposta clara: ou se rejeita ou se aceita a hipótese nula estatística. Rejeitar uma hipótese nula pode ou não ser suficiente para responder à questão de estudo. Pode haver outras explicações possíveis para os resultados e o pesquisador deverá considerar o quanto as explicações alternativas são plausíveis.

PASSO 4

Determinar quais variáveis são importantes para a questão.

As duas variáveis mais relevantes para o experimento são Manejo (Com ou Sem) e o número de espécies. Então, para a categoria “Com” da variável Manejo, há seis objetos ou unidades amostrais, que são as comunidades vegetais, enquanto para a categoria “Sem” também há seis objetos. Portanto, no total, há 12 unidades amostrais, e esse é o tamanho amostral.

PASSO 5

Determinar qual o tipo de cada variável.

O número de espécies em cada uma das 12 unidades amostrais é uma variável quantitativa discreta. Algo que se pode registrar como um número inteiro, pois não há “meia” espécie. A categoria na qual se insere cada uma das seis unidades amostrais de cada área de campo é uma variável nominal, que se registra como um nome (Com ou Sem).

PASSO 6

Desenhar uma amostragem que controle ou aleatorize as variáveis de confusão (os descritores que podem influenciar na resposta biológica e que não são o foco da pesquisa; o tipo de controle desses fatores está ligado ao tipo de pesquisa).

Nesse experimento, só se pode responder se há diferença ou não entre o número médio de espécies nos dois tipos de campos. Se houver diferença, o processo atuando poderá ser qualquer coisa dentro de condições e recursos associados ao manejo. Esse é tipicamente um exemplo de experimento natural ou observacional. No entanto, o experimento indicará, com um grau determinado de incerteza, se há um efeito do manejo sobre o número médio de espécies. Se houver diferença no número médio de espécies vegetais entre as duas áreas, poderão ser levantadas suposições sobre quais processos associados com o manejo podem ser mais importantes para essa diferença.

PASSO 7

Escolher o melhor teste estatístico com base no número de variáveis, nos tipos de variáveis, no ajuste esperado para os pressupostos paramétricos e nas hipóteses a serem testadas.

Como o objetivo é comparar médias de uma variável quantitativa (número de espécies) entre categorias de uma variável nominal com duas categorias (Com e Sem manejo), o teste estatístico apropriado pode ser um teste-*t*, cujo objetivo é comparar médias entre dois grupos. Assim que se sabe quais variáveis estão envolvidas e de que tipo elas são, o número de testes estatísticos possíveis é geralmente limitado a um ou dois (pelo menos para os testes deste livro).

PASSO 8

Realizar, se possível, uma análise do poder do teste para determinar um tamanho amostral adequado para a amostragem.

Uma análise de poder do teste requer: (1) uma estimativa do desvio-padrão do número de espécies, que pode ser encontrado na literatura científica já publicada, ou pode ser calculado a partir de uma coleta piloto; (2) um número para o tamanho do efeito (a diferença entre as duas médias). Nesse experimento, pode se julgar que qualquer diferença no número médio de espécies entre as duas áreas é relevante. Então, deve-se selecionar o máximo número de unidades amostrais dentro do tempo disponível, ou seja, deve-se procurar aumentar ao máximo o tamanho amostral.

PASSO 9

Executar a amostragem – separar os objetos e coletar os descritores em cada um.

O experimento foi realizado e o número de espécies vegetais em cada uma das seis unidades amostrais foi medido nas duas áreas com diferentes manejos com uma técnica de coleta padronizada e os dados foram inseridos em uma planilha de cálculo. Observe, na Tabela 3, o modo de entrada das duas variáveis em uma matriz de dados.

Tabela 3 – Número de espécies vegetais campestres em 12 unidades amostrais com e sem manejo por gado

VARIÁVEL NOMINAL	VARIÁVEL QUANTITATIVA DISCRETA
Com	25
Com	25
Com	26
Com	18
Com	23
Com	21
Sem	18
Sem	9
Sem	7
Sem	8
Sem	8
Sem	8

Fonte: Baldissera *et al.* (2010).

PASSO 10

Examinar os dados para verificar se contemplam os pressupostos do teste estatístico. Se não contemplarem, escolher um teste mais apropriado.

Há várias formas de calcular um teste-*t*. Algumas possuem apenas o pressuposto de que a amostragem foi aleatória. Outros têm pressupostos como uma distribuição de frequências da variável quantitativa com uma forma pré-definida, entre outros. Para testar esses pressupostos, podem ser utilizadas ferramentas gráficas ou analíticas. Se os dados não cumprirem esses pressupostos, em geral, ainda se pode utilizar o teste escolhido com alguma cautela.

PASSO 11

Aplicar o teste estatístico escolhido e interpretar os resultados.

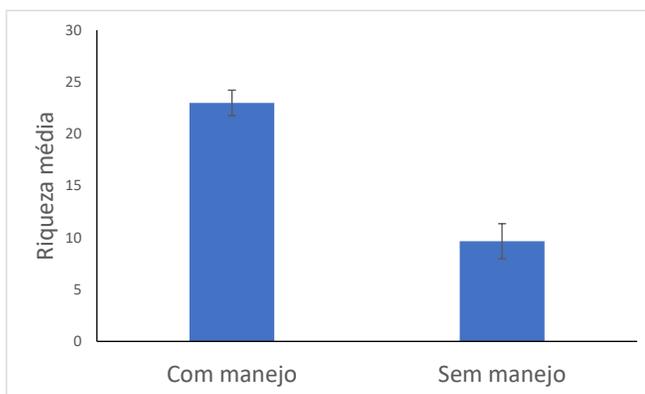
O teste foi realizado utilizando uma planilha eletrônica, uma página da internet, ou um outro programa de computador e o resultado calculou um valor de uma estatística e um valor de $p = 0,00008$. A média de espécies na área com manejo foi 23 e no campo sem manejo 9,67. A interpretação estatística é de que há pouca probabilidade de que a diferença encontrada nas médias deste experimento seja simplesmente uma obra da aleatoriedade. Algo deve estar afetando

essas médias e produzindo a diferença. Nesse caso, um efeito da área Com manejo aumenta o número de espécies vegetais.

PASSO 12

Comunicar os resultados efetivamente, usualmente em forma de gráficos ou tabelas. Utiliza-se, aqui, a estatística descritiva para informar os valores estatísticos que resumem os dados coletados. Para este experimento, uma forma simples de fazer isso é com um gráfico de barras com a variável riqueza de espécies no eixo Y. A altura de cada barra indica o tamanho da média e os desvios são mostrados como linhas com limitadores. A altura de cada linha representa o tamanho da variação dos dados em torno da média.

Figura 4 – Médias \pm desvios-padrões do número de espécies vegetais campestres em campos com manejo e sem manejo



Fonte: Baldissera *et al.* (2010).

Um ponto muito importante é que a execução da amostragem – a coleta de dados – constitui o passo número 9, não o passo 1. Há passos importantes que devem ser decididos antes de realizar a amostragem. Se esses passos forem contemplados, a pesquisa será fácil de entender e de analisar e não será muito grande ou pequena.

2 VARIÁVEIS

Podemos delinear uma variável como uma propriedade que diferencia os objetos ou unidades amostrais do estudo. Os objetos podem ser indivíduos, ou podem ser áreas de uma floresta. Há três tipos principais de variáveis: variáveis quantitativas, que são ordenadas e expressas como números; variáveis nominais, que não são ordenadas e são expressas como nomes; e variáveis ordinais, que são ordenadas semiquantitativas expressas como posições. Deve-se identificar os tipos de variáveis em um experimento para escolher o método correto de análise. Essa identificação deve ser feita antes de realizar o experimento, pois um dos primeiros passos para decidir qual teste estatístico utilizar é determinar que tipos de variáveis serão escolhidas. Quando se sabe quais são as variáveis relevantes, que tipo de variáveis são e quais são as hipóteses do estudo e hipóteses estatísticas, em geral, é relativamente simples escolher o teste.

2.1 VARIÁVEIS QUANTITATIVAS

Variáveis quantitativas são, como o nome indica, descritores que podem ser medidos. Uma medida de uma variável quantitativa é sempre um número. Exemplos incluem comprimento, largura, peso, pH, densidade óssea, número de células, número de indivíduos, número de filhotes, etc.

Essas variáveis podem ser divididas em dois tipos: (1) contínua, tal como o comprimento do corpo de um indivíduo; e (2) discreta, que só assume valores inteiros. As teorias matemáticas embasando os testes estatísticos que envolvem variáveis quantitativas assumem que elas são contínuas. A utilização destes testes estatísticos em variáveis discretas deve ser acompanhada com cuidado, principalmente com respeito ao tipo de distribuição de frequências das medidas realizadas, que, para os testes que veremos, deve se adequar ao que se chama distribuição de probabilidade Gaussiana ou “normal”. Isso quer dizer que há determinadas regras, chamadas de pressupostos, que os dados coletados devem seguir para poderem ser utilizados com confiança em alguns dos testes estatísticos mais usados.

2.2 VARIÁVEIS NOMINAIS

Variáveis nominais classificam objetos em categorias discretas. Exemplos de variáveis nominais incluem sexo (valores possíveis = *machos* e *fêmeas*), genótipo (valores são *AA*, *Aa* ou *aa*), tipo de ecossistema (valores são *Floresta ombrófila*, *Campo*, *Floresta Decidual*, etc.), tratamentos (valores são *Dosagem 1*, *Dosagem 2*, etc.). Uma boa regra é que uma observação individual de uma variável nominal pode ser expressa como uma palavra. Variáveis nominais são, em geral, utilizadas para dividir os objetos do estudo em categorias ou grupos, de forma que outras variáveis possam ser comparadas entre as categorias. Os dois grupos de métodos pós-operatórios discutidos previamente são uma ilustração disso, bem como as duas áreas de campo.

2.3 VARIÁVEIS ORDINAIS

Variáveis ordinais, também chamadas de classificatórias ou de postos, são aquelas para as quais as medidas podem ser colocadas em ordem da menor para a maior, mesmo que os valores exatos sejam desconhecidos. Por exemplo, para saber se homens são mais rápidos do que mulheres em terminar uma prova de análise de dados, se pode listar os indivíduos por ordem de entrega e identificar o sexo. Depois, comparam-se os postos para saber quem é mais rápido.

As variáveis ordinais são importantes porque os testes estatísticos desenhados para esse tipo de variável se baseiam em menos regras do que os testes estatísticos desenhados para as variáveis quantitativas. O que precisa ficar claro é que a alteração do tipo de variável pode fazer com que a hipótese do estudo mude. Então, o melhor é sempre se ater ao tipo de variável medida e ao teste escolhido na fase do projeto de pesquisa.

2.4 CATEGORIZAÇÃO DE VARIÁVEIS

Uma particularidade das variáveis quantitativas discretas e das variáveis ordinais é que elas podem ser convertidas para variáveis nominais quando possuírem poucos valores. Vamos supor que se ministrem três concentrações de uma substância (1 ml, 10 ml e 100 ml) para 30 ratos no laboratório e se meça o peso inicial e final dos bichos. Cada concentração é ministrada para 10 ratos. As

concentrações são números, porém, pode-se tratar cada uma como um nome, por exemplo, Concentração 1, Concentração 10 e Concentração 100. Nesse caso, se categorizou em nomes a variável quantitativa. A decisão sobre como tratar a variável dependerá, em parte, da questão do estudo.

2.5 ESCALA DE LIKERT

A escala de Likert é bastante utilizada em questionários. Apresenta-se uma afirmação como “É importante para todos os biólogos aprender estatística” e se pergunta aos biólogos para escolher entre 1 = discordo fortemente, 2 = discordo, 3 = indiferente, 4 = concordo e 5 = concordo fortemente. Às vezes, se utilizam sete valores e até escalas de 9 a 11 pontos. Questões similares podem ter respostas como 1 = Nunca, 2 = Raramente, 3 = Às vezes, 4 = Frequentemente e 5 = Sempre.

Há muita controvérsia sobre como analisar um item de Likert. Uma opção é tratá-lo como uma variável nominal com tantos grupos quantos forem os valores. Por exemplo, uma categoria seria “Discordo fortemente” e seriam contadas quantas pessoas escolheram essa opção no questionário. Os dados são resumidos pela proporção de pessoas fornecendo cada resposta. Entretanto, isso ignora o fato de que os valores possuem uma ordem de menos concordância para mais concordância, o que é uma informação importante. Para levar em consideração essa informação, pode-se tratar os dados como uma variável ordinal. Tratar os dados como variável quantitativa permite que você os resuma, utilizando medidas como a média e o desvio-padrão, e use um teste estatístico para variáveis quantitativas. Um argumento contra isso é que os dados possuem um pequeno número de valores que provavelmente não cumprirão os pressupostos desses testes.

Um problema para tratar esses dados como variável quantitativa é que as respostas de um valor de Likert são subdivisões grosseiras de uma medida de sentimento e a diferença entre “Discordo fortemente” e “Indiferente” pode não ter o mesmo tamanho da diferença entre “Discordo” e “Indiferente”. Em outras palavras, as respostas não são uma variável com intervalos reais.

2.6 VARIÁVEIS INDEPENDENTES E DEPENDENTES

Outra forma de classificar variáveis é como independentes ou dependentes. Algumas vezes, duas variáveis X e Y são relacionadas de alguma forma. Pode

ser que o valor de Y dependa do valor de X , ou seja, o valor de X de alguma forma explica o valor de Y (MENDENHALL; BEAVER; BEAVER, 2020). A variável que depende é chamada de dependente (resposta) e a variável que explica é chamada de independente (explicativa ou preditora). Por exemplo, besouros tratados com dietas com 10 concentrações diferentes de açúcar podem variar na taxa de crescimento. A concentração de açúcar é a variável explicativa e a taxa de crescimento é a variável resposta. Cada um dos três tipos de variáveis (quantitativa, nominal ou ordinal) pode ser tanto independente, quanto dependente. Por exemplo, para saber se o sexo afeta a temperatura corporal de um animal, o sexo será a variável independente e a temperatura, a variável dependente. Para saber se a temperatura de incubação de ovos afeta o sexo nas tartarugas, a temperatura é a variável independente e o sexo é a variável dependente.

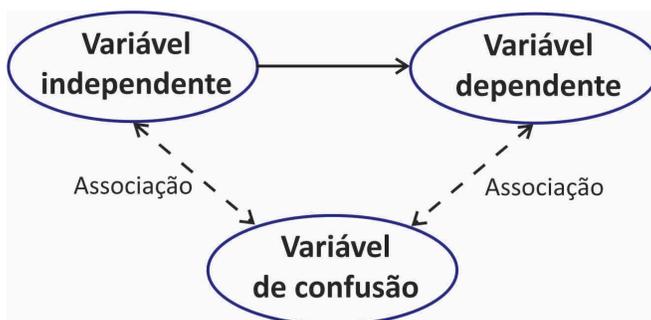
A determinação da variável independente e dependente é meio intuitiva. No exemplo dos besouros, é óbvio que o açúcar é que pode influenciar na taxa de crescimento e não o contrário! A quantidade de açúcar estava sendo manipulada exatamente para gerar um efeito nos ratos. Mas, essa determinação pode ser enganadora, às vezes, principalmente quando se está fazendo um experimento natural ou observacional. Em geral, variáveis nominais são variáveis independentes (mas há exceções). Operacionalmente, é importante saber qual tipo de variável, pois isso pode decidir como vai se analisar os dados. Em outras ocasiões, não há necessidade de decidir se uma variável é independente ou dependente. Por exemplo, se medirmos o conteúdo de nitrogênio do solo e a densidade de plantas dentes-de-leão, se pode pensar que o conteúdo de nitrogênio é uma variável independente e a densidade da planta é dependente. Assim, o conteúdo de nitrogênio afeta onde as plantas vivem. Talvez as plantas utilizem bastante nitrogênio do solo e, então, a densidade da planta deveria ser a variável independente. Ou talvez uma terceira variável que não foi medida, tal como o pH do solo, afete tanto o conteúdo de nitrogênio, quanto a densidade da planta e isso introduz o próximo tópico.

2.7 VARIÁVEIS DE CONFUSÃO

Uma variável de confusão é qualquer descritor que não foi medido, mas que pode afetar o resultado do teste por ter influência sobre as variáveis medidas. Isso pode levar a conclusões errôneas sobre as relações entre a variável resposta e a variável ou variáveis explicativas.

Devido a fatores genéticos, ontogenéticos e ambientais, dois organismos, ou duas amostras de tecidos, ou duas células não são exatamente os mesmos. Isso significa que, quando se desenha um experimento com amostras que diferem na variável explicativa X , as amostras também serão diferentes em outras variáveis que se pode, ou não, estar consciente. Se estas variáveis de confusão afetam a resposta de Y , elas podem levar a pensar que há uma relação entre X e Y quando realmente não há. A Figura 5 ilustra essa associação entre as variáveis.

Figura 5 – Representação da possível influência de uma variável de confusão sobre o comportamento de uma variável resposta



Fonte: Elaborada pelo autor (2021).

Como um exemplo de variável de confusão, pensemos em um experimento no qual o objetivo é saber se o número de filhotes de aranhas varia nas quatro estações do ano. Nesse caso, estação do ano é a variável preditora, ou seja, que explicará o número de filhotes. Se forem contados os números de filhotes de aranhas em uma mata no verão, outono, inverno e primavera, pode-se comparar o número médio deles em cada estação. Porém, as variações climáticas estacionais também afetam outras variáveis que podem interferir na fecundidade. Sabe-se que a qualidade das presas capturadas por uma aranha é importante para seu desenvolvimento ontogenético (TOFT, 2012). Então, talvez o número de presas com mais qualidade varie durante a mudança das estações. A variação no número destas presas pode afetar a fecundidade. Quanto menos presas de qualidade uma aranha come, menor o sucesso em produzir ovos maduros. Então, se houver variação no número de filhotes entre as estações, isso pode ser por causa da diferença no número de presas de qualidade no ambiente. Sem saber disso, a tentação do pesquisador seria afirmar que “a estação do ano explica a fecundidade de aranhas”. Não é uma afirmação errada, mas também não

é 100% correta, pois há outras variáveis que atuam “por trás dos panos” e o número de presas é uma delas.

A elaboração de um desenho amostral que possa minimizar o efeito de variáveis de confusão é criticamente importante. Uma maneira, é controlar uma possível variável de confusão, ou seja, deixá-la idêntica para todos os objetos da análise. Por exemplo, pode-se criar aranhas em um viveiro e simular as alterações climáticas das quatro estações experimentalmente. Cada aranha receberá a mesma alimentação e será disponibilizado o mesmo número de machos para as fêmeas. Então, caso haja diferenças na fecundidade média, a evidência será forte que o fator mais importante é a diferença nas condições climáticas.

O controle de variáveis de confusão é difícil com organismos que vivem fora do laboratório. No laboratório, todas as aranhas, ratos, camundongos, moscas da fruta e peixes zebra terão a mesma idade, sexo e serão da mesma linhagem. Crescerão dentro do mesmo tipo de caixa e comerão a mesma comida. Mas há sempre algumas possíveis variáveis de confusão que não podem ser controladas. Os organismos podem ser todos da mesma linhagem genética, mas novas mutações poderão inserir algumas diferenças entre eles. Pode-se dar a mesma comida e água, mas alguns podem comer ou beber mais que outros. Após controlar todas as variáveis possíveis, é importante lidar com qualquer outra variável de confusão e, para isso, pode-se usar a aleatorização e o pareamento.

2.7.1 Aleatorização

Após desenhar o experimento para controlar quantas variáveis de confusão forem possíveis, se deve aleatorizar as amostras para se certificar de que elas não diferem em outras variáveis de confusão, que não puderam ser controladas. Em resumo, se houver variáveis de confusão, que elas sejam distribuídas aleatoriamente no experimento. Aleatório significa que qualquer objeto do conjunto a ser analisado tem a mesma probabilidade de ser escolhido. Então, aleatorizar significa escolher de forma aleatória, sem preferências. Há alguma confusão nisso, e algumas pessoas pensam que aleatorizar e escolher são coisas diferentes. Na verdade, aleatorizar é um tipo de escolha. O oposto da aleatorização é quando se escolhem os objetos com alguma característica em comum, uma escolha preferencial.

Podemos separar a escolha aleatória da atribuição aleatória. Na primeira, há escolha aleatória de um objeto do universo amostral para compor uma

amostra. A atribuição aleatória se refere a como os objetos são atribuídos a um determinado grupo ou controle experimental. No nosso exemplo sobre os métodos pós-operatórios, as pessoas foram atribuídas a um ou outro grupo.

A maneira mais simples de escolher ou atribuir aleatoriamente é sortear. Por exemplo, para distribuir um bando de ratos em diferentes tratamentos, pode-se rotular os bichos com números, escrever cada número num papel, colocar os números numa sacola e escolher às cegas cada indivíduo, assinalando um tratamento para cada um alternadamente. Ou pode-se jogar um dado cada vez que se pega um rato para decidir qual tratamento administrar (se forem seis tratamentos!).

2.7.2 Pareamento

Às vezes, há muita variação nas variáveis de confusão, o que se torna difícil controlar. Mesmo com a aleatorização, pode haver muita variação na variável resposta que dificultará a detecção de uma diferença causada por uma variável explicativa. Isso é particularmente verdade para seres humanos. Vamos supor que se queira testar se o óleo da planta valeriana é repelente de mosquito. Se ratos forem testados, haverá muitos ratos da mesma idade, sexo e linhagem genética. Desses, metade será escolhida para aplicar o óleo vegetal e será colocada em um lugar cheio de mosquitos por um período e será contado o número de mordidas de mosquitos. Isso seria um bom experimento, bem controlado e, com um número limitado de ratos, seria possível ver se o óleo vegetal causou mesmo uma pequena mudança no número de mordidas. Mas se alguém quisesse testar o óleo vegetal em seres humanos durante seu dia a dia, não teria um conjunto deles com as mesmas características. As pessoas diferem em uma série de características, que não se consegue controlar (desde intrínsecas até o desodorante que a pessoa usa). A grande variação no número de mordidas de mosquitos entre as pessoas significaria que, se o óleo vegetal tivesse um efeito pequeno, seria necessário um número enorme de pessoas para notar uma diferença significativa.

O pareamento, então, é uma maneira de reduzir o ruído devido às variáveis de confusão. Geralmente, se utiliza desse artifício quando a variável explicativa é nominal com dois valores, tais como “droga” vs. “placebo.” Observações são feitas em pares, um para cada valor da variável nominal, que sejam as mais similares possíveis para as variáveis de confusão. Os pares podem ser diferentes partes da mesma pessoa. Por exemplo, pode-se passar o óleo vegetal em um braço da pessoa e o placebo (algo que se sabe que não tem efeito sobre as pica-

das dos mosquitos) no outro braço. A variação no tamanho da diferença entre os dois braços em cada pessoa é muito menor do que a variação entre diferentes pessoas. Então, não é necessário um tamanho amostral tão grande para detectar diferenças na quantidade de mordidas de mosquito entre o óleo vegetal e o placebo. Obviamente, deve se escolher aleatoriamente em qual braço colocar o óleo vegetal a cada observação.

Outra maneira de pareamento inclui experimentos antes e depois. Pode-se contar o número de picadas de mosquito em uma semana, colocar o óleo vegetal nas pessoas e, então, medir novamente para ver se o número de picadas diminuiu. Com esse tipo de experimento, é importante se certificar de que a variável resposta não mudou (talvez o clima tenha mudado e os mosquitos pararam de picar); então, é melhor usar o placebo numa semana e o óleo vegetal na outra e aleatorizar em qual pessoa o óleo vegetal ou o placebo será aplicado primeiro.

Enquanto em experimentos de laboratório parece ser mais fácil realizar pareamentos (controlar as diferenças de categorias), em experimentos de campo, e naturais ou observacionais isso pode ser um pouco mais complicado. Estes experimentos envolvem análises de variáveis que, geralmente, estão expostas às variações naturais. Dessa forma, nesses tipos de experimento, o pesquisador deve ter uma maior preocupação com a escolha das amostras, principalmente no que diz respeito ao processo de aleatorização.

3 DISTRIBUIÇÕES DE PROBABILIDADE

Experimentos biológicos são baseados em variáveis aleatórias, o que significa que os resultados não podem ser pré-determinados. Sempre que se repete o experimento e se olha o resultado, observam-se variações, o que os deixa sem previsibilidade exata. A variável medida é dita aleatória se os valores são determinados por processos acidentais, que não estão sob o controle do observador. Uma distribuição de probabilidade mostra quais as possibilidades de cada resultado para um experimento ou evento. Probabilidade é uma medida da incerteza de fenômenos. Uma distribuição pode ser definida com qualquer variável aleatória, que tem um resultado incerto ou não previsível.

Há muitas distribuições de probabilidades. Ou seja, há várias formas dos valores resultantes de medidas se distribuírem. A distribuição de probabilidade é uma ferramenta que descreve o comportamento de uma variável aleatória. Quando se realiza uma série de medidas, os valores medidos se distribuem ao longo de uma amplitude do menor até o maior valor. Entre esses dois valores extremos, se tem a distribuição dos demais. Se um valor ocorre mais de uma vez durante a medição, ele é mais frequente e se diz que ele tem uma maior probabilidade de ocorrer nessa distribuição. As distribuições de probabilidades são utilizadas, entre outras coisas, para calcular intervalos de confiança e regiões críticas para os testes de hipótese. Geralmente, em estatística, se usa uma letra maiúscula para representar uma variável aleatória e uma letra minúscula para representar um de seus valores. Sendo assim, X representa a variável aleatória; $P(X)$ representa a probabilidade de X ; e $P(X = x)$ se refere à probabilidade de que a variável aleatória X seja igual a um valor x . Por exemplo, $P(X = 1)$ se refere à probabilidade da variável aleatória X ser igual a 1.

Como exemplo, suponhamos que você jogue uma moeda duas vezes. Esse experimento pode ter até quatro resultados possíveis: Cara/Cara, Cara/Coroa, Coroa/Cara e Coroa/Coroa. Vamos fazer com que a variável X represente o número de Caras desse experimento. A variável X pode assumir os valores 0 Cara, 1 Cara e 2 Caras. Uma distribuição de probabilidade é uma tabela ou uma equação que liga cada resultado do experimento com sua probabilidade de ocor-

rência. A Tabela 4 associa cada resultado do experimento da moeda com sua frequência relativa e é uma representação da distribuição de probabilidade da variável aleatória X amostrada no experimento.

Tabela 4 – Distribuição de probabilidades de ocorrer Cara no experimento de jogar uma moeda duas vezes

Número de Caras	Frequência N	Proporção f	Porcentagem $f100$
0	1	0,25	25
1	2	0,50	50
2	1	0,25	25
3	4	1	100

Fonte: Elaborada pelo autor (2021).

A soma da primeira coluna informa o número de eventos bem-sucedidos, ou seja, que contém Cara. A soma da segunda coluna informa o número de eventos nesse experimento. A terceira coluna informa as proporções, que são as probabilidades de ocorrência de “Cara” em cada um dos eventos possíveis no experimento. Observe que o somatório da terceira coluna é 1, que é o limite de probabilidade. Note que, apesar de não saber se um evento de jogar a moeda resultará em Cara ou Coroa, o número de resultados possíveis em cada evento é sempre o mesmo. Esse tipo de experimento é chamado de binomial, pois (1) consiste em n tentativas repetidas; (2) cada tentativa pode resultar em apenas dois possíveis resultados, um “sucesso” ou um “fracasso”; (3) a probabilidade de sucesso, P , é a mesma em cada tentativa; e (4) as tentativas são independentes, ou seja, o resultado de uma tentativa não afeta o resultado de outra tentativa (SOKAL; ROHLF, 2009).

Analisemos um exemplo biológico. A Tabela 5 mostra a distribuição de frequências e porcentagens do número de indivíduos para cada uma das 40 espécies de besouros coletadas em uma área ripária do sul do Brasil.

Tabela 5 – Frequências e porcentagens do número de indivíduos de 40 espécies de besouros

Número de indivíduos	Frequência n_i	Proporção f_i	Porcentagem $f_i 100$
1	16	0,4	40
2	4	0,1	10
3	1	0,025	2,5
4	2	0,05	5
5	1	0,025	2,5
6	1	0,025	2,5
7	1	0,025	2,5
8	2	0,05	5
9	2	0,05	5
10	4	0,1	10
11	5	0,125	12,5
12	1	0,025	2,5
78	40	1	100

Fonte: Elaborada pelo autor (2021).

A soma da primeira coluna informa o número total de indivíduos coletados. A soma da segunda coluna informa o número total de espécies coletadas, enquanto a terceira coluna mostra a proporção de espécies que possuem cada quantidade de indivíduos. Ela informa que o número 1 ocorreu com uma proporção de 0,40, ou 40%. De longe, é a maior proporção. Uma forma de analisar isso é que podemos esperar encontrar um indivíduo de besouro em aproximadamente 40% das espécies, em uma nova coleta com o mesmo tamanho amostral e no mesmo lugar. Outra forma, mais biológica, é inferir que a maior parte das espécies é rara, pois possui apenas um indivíduo. Isso tem uma implicação conservacionista, já que espécies com apenas um indivíduo podem estar mais propensas à extinção local por efeito de algum distúrbio.

Note que a variável da primeira coluna é discreta, assim como a variável Cara da moeda, ou seja, essas variáveis só podem assumir valores inteiros. Quando a variável é discreta, pode-se associar uma probabilidade positiva para cada valor de x e conseguir a distribuição de probabilidade de x . As distribuições de probabilidade que resultam de variáveis discretas são chamadas distribuições de probabilidade discretas. O experimento com a moeda resultará em uma distribuição de probabilidade discreta chamada de binomial. Para o experimento dos besouros, ao contrário do experimento da moeda, não se sabe

de antemão quais os valores que o número de indivíduos assumirá. O tipo de distribuição, em tais casos, é acessado depois da coleta dos dados. Um tipo de distribuição que pode modelar uma variável discreta com valores de $x = 0, 1, 2, 3...$ é a de Poisson (MENDENHALL; BEAVER; BEAVER, 2020). Outra forma seria caracterizá-la como sendo uma distribuição normal.

Mas e se medirmos uma variável quantitativa contínua, como distribuir as frequências e proporções? Analisemos outro exemplo. Dessa vez, olharemos para 24 medidas do coleto de plântulas da espécie *Mimosa bimucronata* (Fabaceae). O coleto é a região de transição entre o caule e a raiz. Os 24 valores são diferentes e a amplitude dos dados variou do menor diâmetro de coleto = 0,89 mm até o maior tamanho de coleto = 2,03 mm. A solução é agrupar os dados por intervalos de diâmetro do coleto, também chamadas classes do coleto.

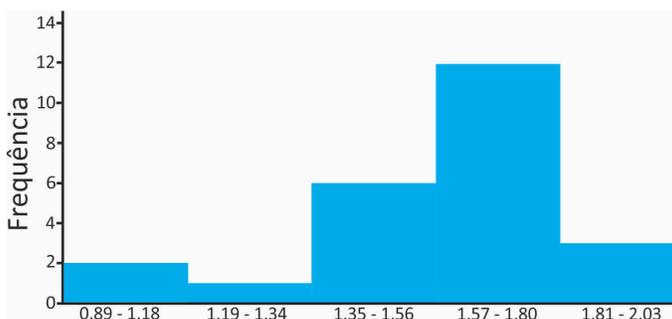
Tabela 6 – Distribuição de frequências e porcentagem das medidas do diâmetro de 24 coletos

Classe	Frequência n_i	Frequência relativa f_i	Porcentagem $f_i 100$
0,89 – 1,18	2	0,083	8,3
1,19 – 1,34	1	0,041	4,1
1,35 – 1,56	6	0,250	25
1,57 – 1,80	12	0,500	50
1,81 – 2,03	3	0,125	12,5
	24	1	100

Fonte: Elaborada pelo autor (2021).

Na Tabela 6, note que a segunda coluna mudou de nome em relação à Tabela 3. Você pode chamar esses valores de proporções ou frequências relativas. Também é bom notar que os dados da variável quantitativa discreta também podem ser tabulados em classes. A vantagem desse tipo de tabulação é que se pode fazer um gráfico chamado histograma, uma ferramenta que mostra como as medidas estão distribuídas em classes. Cada coluna representa uma classe, que define os limites nos quais os valores estão distribuídos.

Figura 6 – Histograma da variável diâmetro do coleto de 24 plântulas



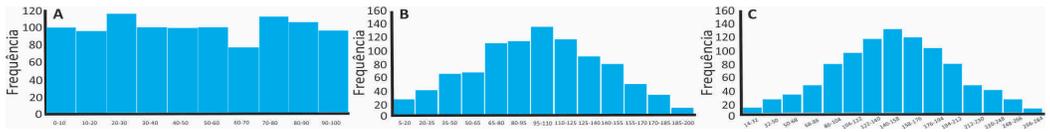
Fonte: Elaborada pelo autor (2021).

A Figura 6 mostra o histograma das medidas do coleto. Note que a maior parte dos diâmetros do coleto está na classe 1,57 – 1,80. São 12 medidas ou 50% do total. Se juntarmos a segunda classe mais frequente, 1,35 – 1,56, serão 18 medidas. Isso dá 75% das medidas. Como vimos anteriormente, as proporções são também probabilidades. O que isso quer dizer? Que, se alguém repetir esse experimento, há 75% de probabilidade dos valores do diâmetro do coleto ocorrerem entre 1,35 e 1,80 mm. Pode-se dizer que estes valores são mais prováveis de ocorrer em uma reamostragem desse universo amostral. Os outros valores são improváveis ou raros. O histograma mostra que há um certo desvio do pico da distribuição para o lado direito. Algumas distribuições podem ter um deslocamento para a esquerda, outras podem ter o pico mais ao centro. Esses diferentes padrões de distribuição de probabilidades é que definem os tipos de distribuição que são utilizados para modelar a distribuição das variáveis aleatórias quantitativas discretas ou contínuas coletadas nos experimentos científicos.

3.1 A DISTRIBUIÇÃO NORMAL

Os testes estatísticos para variáveis quantitativas em geral assumem que a distribuição de probabilidades das observações segue a curva normal ou de Gauss (em forma de sino). Muitas variáveis quantitativas na biologia se distribuem em um padrão normal. A soma de muitas variáveis diferentes com distribuições de valores diferentes segue uma distribuição normal – teorema do limite central (KROKHMAL, 2011). Não importa qual a forma da distribuição das variáveis aleatórias individuais, a soma sempre será normal. A distribuição da soma se ajusta à distribuição normal mais diretamente conforme o número de variáveis aumenta.

Figura 7 – Distribuições de frequências



Fonte: Elaborada pelo autor (2021).

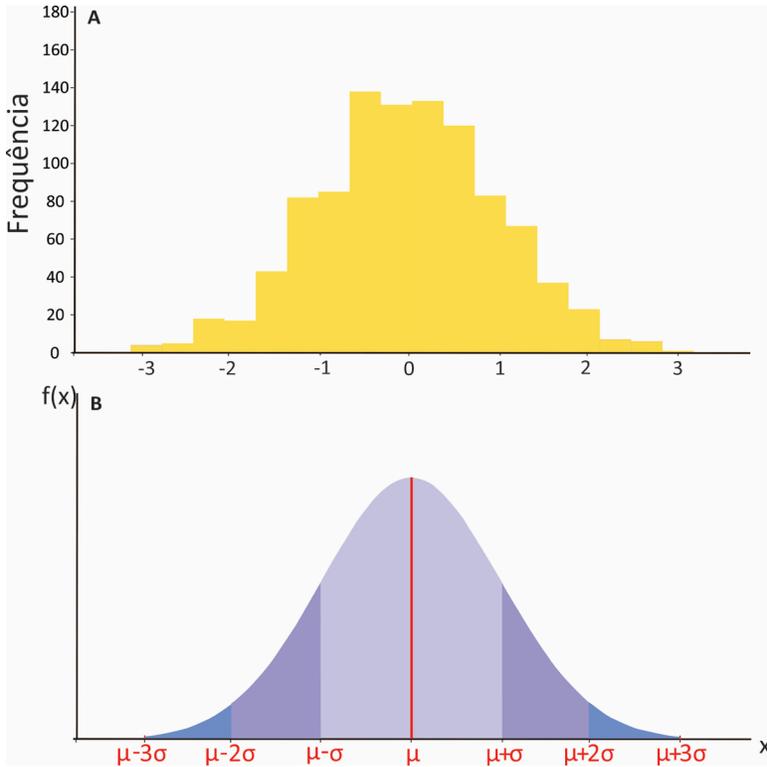
- A: variável aleatória uniforme
- B: soma de duas variáveis aleatórias uniformes
- C: da soma de três variáveis aleatórias uniformes

Os gráficos na Figura 7 são histogramas de frequência com 5000 números. O gráfico A mostra a distribuição de uma variável uniforme entre 0 e 100. A primeira classe mostra os valores entre 0 e 10, a segunda classe os valores entre 10 e 20 e, assim, sucessivamente. Há basicamente a mesma frequência para cada classe como é de se esperar para uma variável uniforme. Isso significa que há basicamente o mesmo número de medidas com os valores entre 1 e 10, entre 10 e 20 e, assim, sucessivamente. Os outros gráficos mostram as distribuições da soma de duas e três variáveis aleatórias com a mesma distribuição uniforme. Note que, quanto mais variáveis uniformes aleatórias são somadas, mais a distribuição de frequências da soma se aproxima de uma curva em forma de sino. Pode-se fazer uma analogia com uma variável biológica, que é o resultado da soma de vários fatores diferentes.

A Figura 8 mostra em (A), uma distribuição amostral de frequências de 1000 números gerados aleatoriamente para se ajustar a uma distribuição normal e em (B), a função de densidade de probabilidade de uma variável aleatória normal padrão com média (μ) = 0 e desvio-padrão (σ) = 1. A distribuição normal possui algumas propriedades que a tornam especialmente úteis para realizar inferências estatísticas. Em primeiro lugar, a média é o valor que localiza o centro da distribuição e a distribuição é simétrica em torno dessa média. Em segundo lugar, o desvio-padrão determina a forma da distribuição. Valores grandes de σ diminuem a altura da curva e aumentam a dispersão, enquanto valores pequenos aumentam a altura e diminuem a dispersão. A probabilidade de ocorrência de algum valor qualquer na distribuição normal padrão é calculada pela área que abrange o valor de interesse até o centro da distribuição. Em torno de 68,3% de todos os valores da distribuição estão entre -1 e +1 desvio-padrão e em torno de 95,4% estão entre -2 e +2 desvios-padrões. Com três

desvios-padrões, a distribuição engloba em torno de 99,7% dos valores. Essa característica implica que valores em torno da média possuem maior probabilidade de ocorrer do que valores mais afastados dela.

Figura 8 – (A) Distribuição de frequências de uma variável aleatória e (B) função densidade de uma variável aleatória normal padrão



Fonte: Elaborada pelo autor (2021).

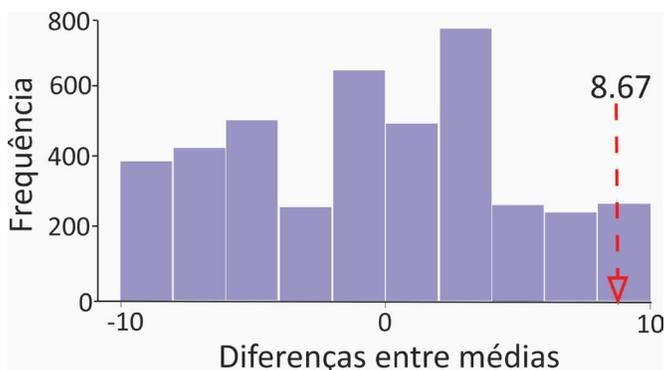
Além de muitas variáveis biológicas possuírem uma distribuição aproximadamente normal e das somas de variáveis não normais também se aproximarem dessa distribuição, muitos testes estatísticos podem ser derivados para a distribuição normal. Quase todos os testes mostrados aqui assumem que foi realizada uma amostragem de uma variável aleatória com distribuição normal. Apesar disso, esses testes funcionam bem, mesmo que a distribuição se afaste da normal. Alguns testes, inclusive, funcionam mesmo que haja grandes desvios da normalidade.

4 TESTE DE HIPÓTESE

Basicamente, há duas perguntas que um pesquisador deve fazer no seu estudo: (1) o quão confiável são os resultados obtidos? e (2) o quão provável é que as diferenças entre os resultados observados e aqueles esperados com base em uma hipótese foram produzidas simplesmente de forma aleatória? (SOKAL; ROHLF, 2009). A primeira questão é respondida com o uso dos limites de confiança para uma amostra. A segunda questão leva ao domínio do teste de hipótese.

A aplicação estatística mais usada na pesquisa biológica é o teste de alguma hipótese científica. Portanto, os métodos de análise estatística são importantes porque os resultados dos experimentos geralmente não são claramente definidos e a estatística contribui para sustentar as decisões em relação a hipóteses alternativas. Nesta lógica, um teste estatístico examina um conjunto de dados e, com base em uma distribuição esperada dos dados (por exemplo, a distribuição normal), leva a uma decisão de aceitar ou rejeitar a hipótese de referência com base nessa distribuição esperada (SOKAL; ROHLF, 2009). Revisitemos a Parte 1.4, que trata da comparação das médias entre dois tipos de pós-operatórios. A Figura 9 mostra um histograma representando a distribuição de frequências das 5040 diferenças entre as médias medidas após realizarmos as permutas dos sete números. Qual seria o valor esperado para que a distribuição dos sete valores originais em cada grupo representasse simplesmente uma aleatoriedade? Veja que a diferença original entre as médias (8,67) está deslocada para uma das caudas da distribuição. A frequência com que esse valor ou valores maiores ocorre não é alta, ou seja, é um evento relativamente raro. Mas o que seria realmente raro? Talvez 5%, 1%, ou 0,1%?

Figura 9 – Distribuição de frequências das diferenças entre 5040 médias calculadas por permutação



Fonte: Elaborada pelo autor (2021).

Então, a lógica dessa abordagem é estimar a probabilidade de obter o resultado observado no experimento, ou algo mais extremo. Se a probabilidade calculada no teste estatístico (o valor de p) é menor do que um valor preestabelecido, então se pode concluir que seja difícil que o resultado seja fruto do acaso. No caso de nosso exemplo, o valor de $p = 0,15$ e pode ser definido como a probabilidade, sob um modelo estatístico proposto, de que um resumo estatístico dos dados seria igual ou maior do que o valor observado (WASSERSTEIN; LAZAR, 2016). Ou seja, se alguém quer comparar se dois grupos são diferentes para uma variável qualquer, o valor de p indicará o quão provável é que a diferença encontrada entre os dois grupos seja grande o suficiente para aceitar que há o efeito de um fator atuando sobre os grupos. A chave é que o valor de p tem a ver com uma probabilidade associada a uma hipótese de referência do pesquisador.

4.1 HIPÓTESES NULA E ALTERNATIVA

A hipótese nula (H_0) é a hipótese de referência do pesquisador. Em geral, a hipótese nula é de que os objetos sejam iguais uns aos outros, como no exemplo dos grupos de pós-operatório; ou que são independentes entre si; ou não correlacionados uns aos outros; ou iguais a uma expectativa teórica. Por exemplo, se alguém medir o tamanho de patas de cães da raça pastor alemão adultos machos e fêmeas, a hipótese nula é que a média de comprimento da pata em machos é igual à média de comprimento da pata em fêmeas. Por outro lado, se alguém contar o número de machos e fêmeas nascidos de um conjunto

de ninhadas, a hipótese nula pode ser que a razão entre machos e fêmeas nos filhotes é igual a uma expectativa teórica de 1:1 (ou seja, 50% serão machos e 50% serão fêmeas). A hipótese alternativa (H_1) é a que será escolhida se a hipótese nula desviar do esperado. Por exemplo, uma hipótese alternativa é a de que cães pastores-alemães machos possuem comprimento médio de pata diferente das fêmeas; outra é a de que a razão sexual é diferente de 1:1.

Quando se pensa em uma variável aleatória quantitativa dividida em dois grupos, é meio óbvio que há alguma diferença entre as duas médias, mesmo que seja minúscula. Por exemplo, a média do comprimento das patas de machos é 7,13 cm e de fêmeas é 7,03 cm. A razão sexual teórica de 1:1 também pode ser afetada por muitas outras variáveis e a ninhada pode ter um desbalanço no número de machos e fêmeas. Dessa forma, o importante, nesse caso, é saber se a diferença é grande o suficiente para rejeitar a hipótese nula, que é a referência do pesquisador. No caso das patas de cães pastores alemães, a diferença de 0,10 cm é grande o suficiente para se inferir que machos têm patas maiores? Ou seja, o teste de hipótese procura saber “Qual a probabilidade de conseguir um desvio da hipótese nula que seja tão extremo se essa hipótese for realmente verdade?” No nosso exemplo dos métodos de pós-operatório, a probabilidade é 0,15. Somente quando esta probabilidade (valor de p) for suficientemente baixa, pode-se rejeitar a hipótese nula. O suficientemente baixa é determinado pelo pesquisador. Então, o objetivo do teste de hipótese estatística é estimar a probabilidade de conseguir os resultados observados sob a hipótese nula. No nosso exemplo, sob a hipótese nula de que a diferença entre os dois métodos pós-operatórios é um evento imprevisível de ocorrer. Quanto maior essa probabilidade (maior valor de p), maior a evidência de que a hipótese nula é a verdadeira naquele teste. Inversamente, quanto menor essa probabilidade, maior a evidência de que a hipótese alternativa é a verdadeira para o teste.

4.2 HIPÓTESES ESTATÍSTICAS VS. HIPÓTESES DE ESTUDO

É importante distinguir entre hipóteses de estudo e hipóteses estatísticas. “A seleção sexual fez com que os machos de cães pastores alemães evoluíssem patas mais compridas do que as fêmeas” é uma hipótese de estudo, pois declara algo sobre processos evolutivos, nesse caso, seleção sexual do comprimento das patas. “Cães machos possuem comprimentos médios de patas maiores do que fêmeas” é uma hipótese estatística, pois declara algo sobre números, mas nada

sobre o que faz com que os números sejam diferentes. É o pesquisador que decide pela hipótese de estudo, geralmente, com base em uma expectativa teórica ou em outros estudos empíricos. Por exemplo, a hipótese de estudo pode declarar especificamente que um ou outro sexo de cães pastores alemães terá patas mais compridas, mas a hipótese estatística simplesmente declara que há ou não uma diferença nas médias. Neste caso, a hipótese estatística não distingue se as patas dos machos ou das fêmeas são mais compridas. É quem faz a análise que deve declarar em que sentido estará a maior média. Essa é a previsão baseada na hipótese de estudo. O teste estatístico confirmará ou rejeitará essa previsão.

A hipótese de estudo é a primeira na qual se deve pensar, pois é originária de um questionamento interessante sobre a biologia do sistema de estudo. Por exemplo, “O que afeta o comprimento de patas em cães?”. A hipótese nula e alternativa estatística são afirmações sobre os dados que devem se originar da hipótese de estudo: se a seleção sexual favorece o maior comprimento de patas nos machos (uma hipótese do estudo), então a média de tamanho das patas em machos será maior do que em fêmeas (uma hipótese estatística). Quando se rejeita a hipótese de referência estatística, talvez não se tenha evidência suficiente para se rejeitar a hipótese de estudo. Por exemplo, se não há diferença significativa no comprimento da pata entre machos e fêmeas, se poderia concluir que “Não há evidência significativa que a seleção sexual causou cães machos de possuírem patas mais compridas.” Se houver uma diferença estatística significativa no comprimento das patas, talvez isso não seja o suficiente para concluir que o processo seleção sexual causou essa diferença. Pode ser que os machos adultos escolhidos tenham comido mais durante a juventude ou que as patas mais compridas ocorrem porque os machos se movimentam mais e o exercício torna suas patas mais musculosas.

4.3 TESTANDO A HIPÓTESE NULA

O objetivo principal do teste estatístico é determinar se um conjunto de dados observados é diferente do que se poderia esperar sob a hipótese nula, de forma que se possa rejeitar essa hipótese. Consideremos o seguinte exemplo (MCDONALD, 2014): suponha que alguém alimente muitos frangos com chocolate e, então, olhe a razão sexual das crias. Se, em uma ninhada de 48 filhotes, forem encontradas 25 fêmeas e 23 machos, com certeza há mais fêmeas que machos, mas a razão sexual é praticamente a esperada de 1:1 (em torno de 1,042:0,96). Não há razão para supor que o chocolate influencia na razão se-

xual e nem para rejeitar a hipótese nula de uma razão sexual de 1:1 de machos e fêmeas. Se houver 47 fêmeas e 1 macho, pode-se dizer que esse resultado é extremamente raro, se a hipótese nula de uma razão sexual 1:1 é verdadeira. Ou seja, não se esperaria esse resultado, uma vez que se sabe que a razão sexual de frangos é 1:1. Então, nesse caso, se rejeitaria a hipótese nula e aceitaríamos a hipótese alternativa de que o chocolate mudou a razão sexual. Porém, e se houver 33 fêmeas e 15 machos? Nitidamente houve um desvio da razão 1:1. Entretanto, isso não é suficiente para afirmar categoricamente que o chocolate afeta o sexo de frangos. Note que os resultados podem mudar de experimento para experimento e, apesar de raro, um desvio é perfeitamente aceitável probabilisticamente falando. Então, deve se perguntar “Qual a probabilidade de conseguir um desvio da hipótese nula que seja igual ou maior do que esse, aleatoriamente, se a hipótese nula for realmente verdade?” Somente quando esta probabilidade (valor de p) estiver contida em uma amplitude predeterminada de valores de p se pode rejeitar a hipótese nula.

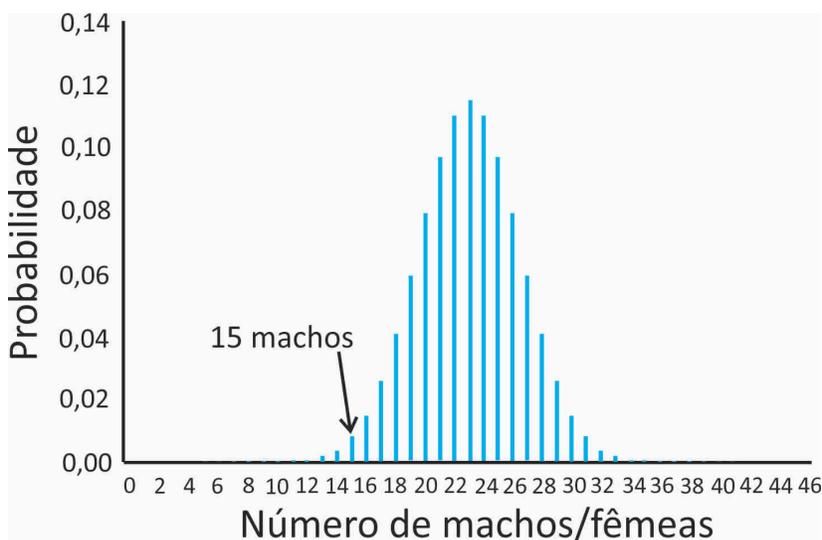
Então, o resultado de 33 fêmeas e 15 machos é realmente tão raro assim de ocorrer para que possamos rejeitar a hipótese nula? Para responder a essa pergunta, é necessário calcular a probabilidade de se conseguir um desvio tão grande devido somente ao acaso. Usaremos, então, esse novo exemplo para aprofundar a análise que realizamos no experimento dos métodos pós-operatórios.

4.4 VALOR DE p

A Figura 10 mostra as probabilidades de se conseguir cada número possível de machos ou fêmeas, de 1 até 48, sob a hipótese nula de que 0,5 dos indivíduos são machos ou fêmeas (razão sexual 1:1). Note que, como o experimento da moeda, esse só possui dois resultados possíveis. Como se pode observar, a probabilidade de se conseguir 15 machos de um total de 48 frangos é em torno de 0,004. Parece que é uma probabilidade bem pequena, não é mesmo? Entretanto, essa é a probabilidade de se conseguir exatamente 15 machos. O que queremos saber é a probabilidade de conseguir 15 ou menos machos. Se aceitarmos 15 machos como evidência de que a razão sexual está distorcida, também temos de aceitar 14, 13, 12... 0 machos como evidência dessa tendência distorcida na razão sexual. Portanto, temos de somar as probabilidades de todos esses resultados (15 até 0). A probabilidade de conseguir 15 ou menos machos de 48 indivíduos sob a hipótese nula é mais ou menos 0,0066. Isso significa que, se houver um in-

finito número de frangos, metade machos e metade fêmeas (razão 1:1) e alguém pegar um conjunto de amostras aleatórias de 48 frangos, 0,6% das amostras teriam 15 ou menos frangos machos. Se houver “infinitos” frangos (vamos dizer 1.000.000), pode-se pegar 100 amostras de 48 frangos. Dessas 100 amostras, em apenas 0,6% delas se esperaria encontrar somente 15 machos.

Figura 10 – Distribuição de frequências das probabilidades de se encontrar machos e fêmeas em uma ninhada de 48 filhotes



Fonte: Elaborada pelo autor (2021).

Esse número – 0,0066 – é o valor de p desse novo experimento. Então “ $p = 0,0066$ ” é uma maneira abreviada de dizer que “A probabilidade de se conseguir 15 ou menos frangos machos de um total de 48 indivíduos *SE* a hipótese nula de que 50% dos frangos são machos é verdadeira é de 0,0066.” Neste exemplo, o interesse era saber qual a probabilidade de se conseguir 15 ou menos machos. Porém, invertendo a lógica, pode-se pensar que a probabilidade de se conseguir 33 ou mais machos também é igualmente improvável.

4.5 O VALOR DE p NO BANCO DOS RÉUS

Se você entendeu a lógica do valor de p associado com um teste de hipótese até aqui, percebeu que essa abordagem é poderosa, pois oferece ao pesqui-

sador uma ferramenta para aceitar ou rejeitar um resultado associado a uma ideia. Seria revolucionário se o chocolate alterasse a razão sexual de frangos. Fêmeas são mais valiosas do que machos. Por isso, ao mesmo tempo que é uma ferramenta útil, pode ser mal-usada e interpretada de forma acidental, ou até mesmo proposital. Por isso, nos últimos anos, tem havido muito debate a respeito da prática do uso de abordagens estatísticas que se utilizam do valor de p . O fato é que o valor de p , apesar de parecer, não é tão confiável, nem tão objetivo quanto parece (NUZZO, 2014). Uma das características mais marcantes que se deve ter em conta quando se realiza um experimento e se utiliza um valor de p para estudar a inferência estatística é que esse valor de p foi obtido *no experimento realizado*. A confiabilidade de que o mesmo valor de p , ou algum valor menor, seja encontrado em outros experimentos é limitada.

A situação de confusão criada pelo uso e abuso do valor de p nas diferentes áreas do conhecimento levou a Associação Americana de Estatística a fazer uma declaração oficial sobre valores de p e significância estatística (WASSERSTEIN; LAZAR, 2016). No documento publicado, a entidade esclarece alguns pontos de confusão, que são resumidos a seguir:

1. Valores de p podem indicar quanto os dados são incompatíveis com um modelo estatístico específico. Um valor de p é uma abordagem para resumir a incompatibilidade entre um conjunto particular de dados e um modelo proposto para os dados. O contexto mais comum é que o modelo esteja ligado a uma hipótese nula, a qual, frequentemente, postula a falta de um efeito, tal como a falta de diferença entre dois grupos. Essa incompatibilidade pode ser interpretada como uma dúvida ou uma evidência contra a hipótese nula ou sobre os pressupostos das análises.

2. Valores de p não medem a probabilidade de que a hipótese estudada é verdadeira, ou a probabilidade de que os dados foram produzidos simplesmente de forma imprevisível. Os pesquisadores não podem tornar um valor de p em uma declaração sobre a verdade de alguma hipótese, ou sobre a probabilidade de que um processo aleatório produziu os dados coletados. O valor de p é uma declaração sobre os dados em relação a uma explicação hipotética específica e não é uma declaração sobre a própria explicação.

3. Conclusões científicas e decisões de negócio ou de políticas não devem ser baseadas exclusivamente no fato de um valor de p estar fora de um determinado limite. Práticas que reduzem a análise de dados a regras inequívocas (tal

como “ $p < 0,05$ ”) para justificar conclusões podem levar a crenças errôneas e tomadas de decisões pobres. Os pesquisadores devem contextualizar vários outros fatores para derivar inferências científicas, incluindo o desenho amostral de um estudo, a qualidade das medidas, a evidência externa para o fenômeno sendo estudado e a validade de pressupostos que embasam a análise. Em resumo, valores de p sozinhos não podem certificar que uma decisão é correta ou incorreta.

4. Uma inferência apropriada exige completa informação e transparência. Conduzir uma análise de dados e informar somente certos valores de p faz com que não se possa interpretar o estudo. Nesse tipo de prática, tipicamente, os pesquisadores “escolhem” valores de p que confirmem alguma hipótese desejada. Essa prática, também chamada dragagem de dados, caça à significância, busca da significância e inferência seletiva, leva a falsos excessos de resultados significativos na literatura científica e devem ser evitados. Conclusões científicas válidas baseadas em valores de p e estatísticas associadas não devem ser tomadas sem ao menos saber quantas e quais análises foram feitas, e como essas análises foram selecionadas e informadas.

5. Um valor de p , ou uma significância estatística, não mede o tamanho de um efeito ou a importância de um resultado. Significância estatística não é a mesma coisa que significância científica, humana ou econômica. Valores pequenos de p não implicam, necessariamente, a presença de efeitos maiores ou mais importantes, bem como valores grandes de p não implicam uma falta de importância ou até mesmo uma falta de efeito.

6. Por si só, um valor de p não fornece uma boa medida de evidência com respeito a um modelo ou hipótese. Um valor de p próximo a 0,05, tomado por si só, fornece apenas uma fraca evidência contra a hipótese nula. Da mesma forma, um valor relativamente grande de p não implica uma evidência a favor da hipótese nula; muitas outras hipóteses podem ser iguais ou mais consistentes com os dados observados.

A conclusão do documento é a seguinte:

A boa prática da estatística, como um componente essencial da boa prática científica, enfatiza [1] princípios de bom desenho amostral e condução, [2] uma variedade de resumos numéricos e gráficos dos dados, [3] entendimento do fenômeno em estudo, [4] interpretação contextualizada dos resultados, [5] divulgação completa, [6] e entendimento lógico e quantitativo correto do que o resumo dos dados significa (WASSERSTEIN; LAZAR, 2016).

4.6 FALSOS POSITIVOS E NEGATIVOS

Depois de realizar um teste estatístico, pode-se rejeitar ou aceitar a hipótese nula. Rejeitar a hipótese nula leva à inferência de que há evidência suficiente de que essa hipótese é incorreta para o experimento em questão. No exemplo dos frangos, a proporção verdadeira de frangos machos, após dar chocolate para um infinito número de fêmeas prenhas, seria menor do que 50%. Dessa forma, pode-se falar em 50% (24 frangos machos) como o resultado esperado e 31,25% (15 machos) como o resultado observado.

Quando se rejeita a hipótese nula, sempre há uma chance de que um erro esteja sendo cometido. A probabilidade (valor de p) encontrada anteriormente (0,0066) é bem baixa, mas não é zero! Ou seja, ainda há uma pequeníssima probabilidade de a hipótese nula ser correta. Então, a hipótese nula pode realmente ser correta e pode ser que os resultados dos experimentos desviem dela simplesmente porque essa é uma das probabilidades possíveis. Então, em uma amostra de 48 frangos, é possível conseguir 15 machos. É até possível, porém muito improvável, de se conseguir zero machos, pois essa é uma das probabilidades. Portanto, sempre há uma possibilidade, embora minúscula, de que os dados enganem o pesquisador e se desviem da hipótese nula. Falso positivo é quando os dados enganam e se rejeita a hipótese nula, mesmo ela sendo a correta, ou, em linguagem estatística, é o erro Tipo I.

Outra maneira que os dados podem enganar é quando se aceita a hipótese nula, mesmo que ela não seja correta. Se a proporção real de frangos fêmeas é 51%, a hipótese nula de uma proporção de 50% não é verdadeira, mas é improvável que se consiga uma diferença significativa da hipótese nula, a não ser que se tenha um tamanho amostral enorme. Falhar em rejeitar a hipótese nula, mesmo que ela seja falsa, é um falso negativo ou um erro Tipo II. Por isso, nunca se diz que os dados mostram que a hipótese nula é verdadeira. Tudo que se pode dizer é que há uma evidência forte o suficiente para aceitá-la.

4.7 NÍVEIS DE SIGNIFICÂNCIA

O nível de significância estatística é a probabilidade de rejeitar a hipótese nula quando ela é verdadeira, ou seja, $P(\text{Erro Tipo I}) = \alpha$. É uma medida da força da evidência que os dados devem apresentar antes de rejeitar a hipótese nula e concluir que há um efeito significativo. Por exemplo, um nível de significância de

0,05 indica que há um risco de 5% em concluir que existe um efeito quando, na verdade, ele não existe, ou seja, $P(\text{Erro Tipo I}) = 0,05$. Níveis de significância menores indicam que é necessária uma evidência mais forte antes de rejeitar a hipótese nula. A convenção na maioria das pesquisas biológicas é utilizar um nível de significância de 5%. Depois que se escolhe o nível de significância antes de fazer a análise, precisa-se comparar o valor de p observado no resultado da análise com ele. Então, para um nível de significância de 0,05, o valor de p da estatística deve ser $< 0,05$ (se lê “menor que”) para rejeitar a hipótese nula. Se p é igual ou maior do que 0,05, não há evidência suficiente para rejeitar a hipótese nula. O nível de significância a ser utilizado depende do custo dos diferentes tipos de erros.

Para o experimento dos métodos pós-cirúrgicos, a média calculada e valores mais extremos tiveram uma probabilidade de 0,15 com base na distribuição de probabilidades das 5040 diferenças. Nesse caso, com um valor de p predefinido de 0,05, devemos aceitar a hipótese nula de que não há diferença entre as duas médias dos dias de recuperação dos métodos pós-operatórios. Isso equivale a dizer que o tratamento não alterou o número médio de dias de recuperação com suficiente força para aceitarmos que ele é realmente diferente do controle. Note que não dizemos que “o tratamento é igual ao controle”. Não é. Há uma diferença. Porém, *para esse experimento*, a diferença não foi grande o suficiente para que pudéssemos ter confiança de que tudo não passou de um artifício dos dados.

Pode-se usar níveis de significância diferentes. Se for usado um nível de significância maior do que os 0,05 convencionais, tal como 0,10, há maior chance de um falso positivo (portanto, aumentando a chance de errar!), mas também se diminui a chance de encontrar um falso negativo (aumentando a chance de detectar um efeito sutil). Caso se use um nível de significância menor do que os convencionais 0,05, tal como 0,01, há menor chance de encontrar um falso positivo (pois, se assumirá que só 1% dos resultados possíveis fornecerão esse falso positivo), mas também será menos provável que se detecte um desvio real da hipótese nula se ele existe.

Deve-se escolher o nível de significância antes de realizar as análises. Não há uma regra específica sobre qual nível de significância é indicado. Porém, é importante pensar nos dois tipos de erros (I e II). Enquanto α é a probabilidade de incorrer no erro Tipo I, β é a probabilidade de incorrer no erro Tipo II. A diferença $1 - \beta$ é o poder do teste estatístico, definido como a probabilidade de rejeitar H_0 quando essa hipótese é realmente falsa. Uma característica importante dessa relação é que, para um tamanho amostral constante, há uma

variação na relação entre α e β . A única forma de diminuir os dois tipos de erros é aumentando o tamanho amostral. Por exemplo, para termos um poder $(1 - \beta)$ de 0,996 com um $\alpha = 0,05$ e um $\beta = 0,004$, o tamanho amostral deve ser $n = 300$ (KIM, 2015). Esse tamanho amostral é impraticável em muitos experimentos. Geralmente, o pesquisador estabelece um $\alpha = 0,05$ por conveniência. Porém, com um tamanho amostral de 10 objetos, por exemplo, a probabilidade de aceitar H_0 , quando ela é falsa, é 80%, considerando uma variável aleatória com distribuição normal. Em experimentos exploratórios ou naturais/observacionais, pode ser difícil definir os parâmetros necessários para calcular o tamanho amostral “ideal”, já que não se sabe a priori uma diferença entre médias e um desvio-padrão esperado, ou um coeficiente de regressão e um intervalo de confiança esperado. Para experimentos em que se podem acessar parâmetros esperados confiáveis do universo amostral, sugere-se fazer um teste de poder para determinar o tamanho amostral, principalmente quando o resultado pode colocar em risco a vida, ou o bem-estar. Para isso, também deve se levar em consideração o que pode se perder ao escolher um valor de α e de β (KIM, 2015). Ao longo deste livro, utilizaremos o nível de significância de 5% ($p < 0,05$) como critério de evidência para rejeitar a hipótese nula em todas as análises. Como a maioria dos tamanhos amostrais dos exemplos é menor do que 30 objetos, isso significa aceitar a probabilidade de mais ou menos 0,8 de estar aceitando H_0 , quando ela é falsa.

4.8 PROBABILIDADES DE UMA CAUDA E DE DUAS CAUDAS

A probabilidade que foi calculada no exemplo dos frangos (0,0066) é a probabilidade de conseguir 15 ou menos machos de uma ninhada de 48 frangos, considerando que a razão sexual é 1:1. Ela é significativa, utilizando o critério de $p < 0,05$ convencional. Entretanto, e quanto à probabilidade de se conseguir 33 ou mais machos? Se a hipótese nula é “A proporção de machos é 0,5 ou mais” e a hipótese alternativa é “A proporção de machos é menor do que 0,5”, então se utiliza o valor de $p = 0,0066$ encontrado. Esse tipo de probabilidade é chamado de uma cauda, pois as probabilidades estão sendo somadas somente em um lado da distribuição mostrada (no caso, do lado esquerdo). Entretanto, se a hipótese nula é “A proporção de machos é 0,5”, então a hipótese alternativa é “A proporção de machos é diferente de 0,5”. Nesse caso, se deve somar a probabilidade de

conseguir 33 ou mais machos à probabilidade encontrada. Esse tipo de probabilidade é chamado de duas caudas. Fazendo isso com o exemplo dos frangos (15 ou menos machos = 0,006 + 33 ou mais machos = 0,006) o valor de $p = 0,012$.

Deve-se decidir que tipo de probabilidade utilizar antes de analisar os dados. A probabilidade de uma cauda deve ser utilizada se, verdadeiramente, se tem uma predição firme sobre em que sentido o desvio acontecerá. Isso tem a ver com a hipótese de estudo. Cada vez que um desvio em qualquer direção seja importante, deve-se usar a probabilidade de duas caudas. Adicionalmente, as pessoas são céticas de probabilidades de uma cauda, especialmente se uma probabilidade de uma cauda é significativa e uma probabilidade de duas caudas com os mesmos dados não é significativa. Se o pesquisador não apresentar uma explicação muito convincente, as pessoas podem pensar que se decidiu usar uma probabilidade de uma cauda após perceber que a probabilidade de duas caudas não foi significativa. Ao longo do livro, trabalharemos com probabilidades de duas caudas.

4.9 APRESENTANDO OS RESULTADOS

Após a análise dos dados, deve-se concluir que há evidência suficiente para rejeitar a hipótese nula e reportar os gráficos, tabelas e valores de p , juntamente com uma medida de dispersão dos dados. Atualmente, todos os programas estatísticos fornecem o valor exato de p de um teste estatístico e é assim que se deve reportar os resultados. É diferente encontrar para duas análises baseadas em duas amostras do mesmo universo amostral ou população estatística um valor de $p = 0,01$ e outro de $p = 0,045$.

Quando os valores de p ficam muito pequenos, pode-se simplesmente reportar " $p < 0,0001$ " ou outro número impressionantemente pequeno. Uma boa prática, inclusive endossada pela Associação Americana de Estatística (WASSERSTEIN; LAZAR, 2016), é informar quais os dados utilizados, como foram coletados, como os testes estatísticos foram escolhidos e realizados e as hipóteses de estudo levantadas, juntamente com o valor do teste estatístico e os graus de liberdade, caso qualquer pessoa queira calcular o valor exato de p . Números podem ser facilmente manipulados, por isso é sempre necessário ser transparente. Lembrando que não há certo ou errado no resultado de uma análise qualquer.

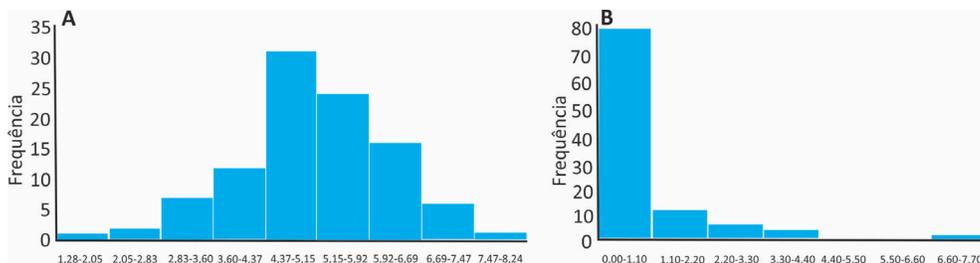
5 ESTATÍSTICA DESCRITIVA DE VARIÁVEIS QUANTITATIVAS

5.1 HISTOGRAMA

O histograma é uma ferramenta que mostra como as medidas estão distribuídas em classes. Já vimos algumas características desse gráfico na parte de distribuição de probabilidades. Essa ferramenta é utilizada com variáveis contínuas, mas variáveis discretas também podem ser utilizadas. Para um conjunto de dados pequenos, um histograma de variáveis discretas pode não fazer muito sentido e o melhor é mostrar a distribuição de cada um dos valores discretos, assim como fizemos na Tabela 3.

O histograma é um gráfico que pode ser utilizado para fazer o diagnóstico do tipo de distribuição da variável aleatória coletada no experimento. A Figura 11 mostra as distribuições de frequências para dois conjuntos de dados. O gráfico A é de uma variável aleatória com distribuição normal e o gráfico B de uma variável aleatória com distribuição chamada de qui-quadrado. Note que, no gráfico A, a distribuição de frequências segue muito bem o modelo de distribuição normal em forma de sino. Já se você fizer um histograma e encontrar um padrão como o do gráfico B, provavelmente não poderá utilizar os testes estatísticos que se baseiam na distribuição normal. Porém, há algumas transformações que se podem utilizar para que a distribuição de uma variável aleatória fique mais parecida com a normal.

Figura 11 – Distribuições de frequências

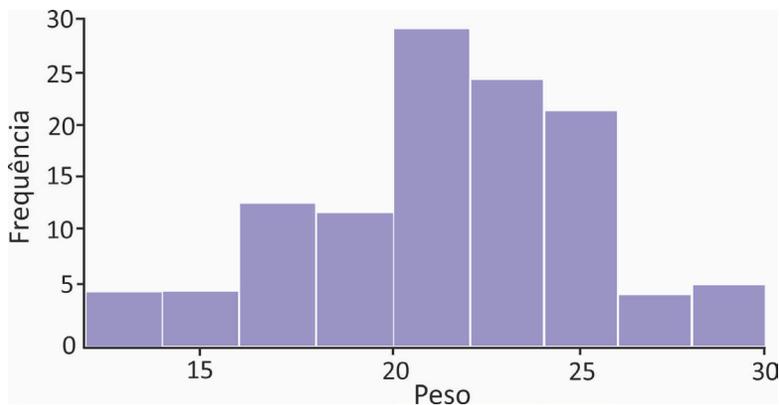


Fonte: Elaborada pelo autor (2021).

A: variável aleatória com distribuição normal
 B: variável aleatória com distribuição qui-quadrado

A Figura 12 mostra o histograma do peso de 128 morcegos da espécie *Sturnira lilium* coletados em um fragmento florestal (CORÁ; OLIVEIRA; GALIANO, 2018). Note que o gráfico mostra um padrão de distribuição parecido com a distribuição normal. Há um pico no centro da distribuição e uma diminuição das frequências para os dois lados. À primeira vista, pode-se dizer que seja uma variável com distribuição normal. Neste caso, podemos dizer, com alguma variância, que a média da variável peso dessa amostra é uma estimativa da distribuição da população dessa espécie no fragmento, que é o universo amostral. Em adição à análise gráfica, podemos realizar o teste estatístico de Shapiro-Wilk, que será apresentado mais adiante.

Figura 12 – Distribuição de frequências do peso de *Sturnira lilium*



Fonte: Elaborada pelo autor (2021).

5.2 MEDIDAS DE TENDÊNCIA CENTRAL

5.2.1 Média aritmética

A média aritmética é a soma das observações dividida pelo número de observações. É a estatística mais comum de tendência central. É simbolizada geralmente colocando uma barra sobre a letra que representa a amostra. Para um conjunto de medidas $x = (1,2,3,4,5)$ onde $x_1 \dots x_n =$ conjunto de valores de x e $n =$ quantidade de valores, a média de x é calculada como:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Ou $1+2+3+4+5 / 5 = 3$. A média funciona bem para valores que se ajustam à distribuição normal. É sensível a valores extremos, que a tornam inadequada para dados que são fortemente distorcidos. Por exemplo, em uma área, 99% das árvores de uma espécie são jovens com 1 m de altura e apenas 1% delas possui mais de 20 metros. Se uma amostra de 20 árvores fosse feita e apenas uma das árvores grandes fosse amostrada, a média de altura seria 2,45 m. Uma amostra que não incluísse uma árvore grande teria uma média de 1 m. A média de uma amostra variará bastante, dependendo se uma árvore grande for amostrada ou não.

Para os dados dos morcegos, a média da variável peso é 21,9 g.

5.2.2 Mediana

Quando se ordena os valores de uma variável do menor para o maior, a mediana é o valor da lista que está no meio. Para um valor ímpar de valores, a mediana é o único valor que está no meio da lista ordenada. Para um valor par, a mediana é a média aritmética dos dois valores que estão no meio dos valores ordenados. Então, para um conjunto com valores x_1, x_2, x_3, x_4 e x_5 , a mediana será x_3 , para uma variável com seis valores, a mediana será a média aritmética de x_3 e x_4 . Então, a mediana do conjunto $A = (1,2,3,4,5)$ é 3. A mediana do conjunto $B = (1,2,4,5)$ é a média de (2,4), que é 3. No exemplo dos morcegos, a mediana dos pesos é 22 g.

5.2.3 Moda

A moda é o valor mais comum em um conjunto de dados. Porém, sua detecção depende do tipo de variável quantitativa. Com variáveis contínuas, dependendo da precisão da medida, pode ser complicado ver a moda. Se forem medidas as alturas de 25 pessoas ao milímetro mais próximo, provavelmente, todas elas terão valores diferentes e não haverá moda. Porém, se medir ao centímetro mais próximo, haverá pessoas com as mesmas alturas.

Raramente, determinar a moda de uma distribuição é útil, mas é útil distinguir entre distribuições unimodais, bimodais, etc. Essa característica determina se a frequência paramétrica de uma distribuição baseada em um conjunto de observações tem um pico, dois picos, etc. Para os dados dos morcegos, a moda do peso é 21 g.

Note que a média, a mediana e a moda do peso dos morcegos são similares. Essa é uma das características da distribuição normal, essas três estatísticas andam juntas.

5.3 MEDIDAS DE DISPERSÃO

Além da medida de tendência central, é desejável ter uma medida de dispersão dos números, conhecida como estatística de dispersão. Adicionalmente, além da média, o desvio-padrão, que é uma medida de dispersão, são os dois descritores de uma distribuição normal. Utiliza-se uma estatística de dispersão para ter um número único que descreve o quão compacto ou espalhado um conjunto de observações se apresenta.

5.3.1 Amplitude

É simplesmente a diferença entre o maior e o menor valor. A amplitude não é muito informativa para propósitos estatísticos, ela depende apenas dos valores maiores e menores e, portanto, dois conjuntos de dados com distribuições muito diferentes podem apresentar a mesma amplitude. Por exemplo, a variável $A = (1,2,5,10)$ tem a mesma amplitude $= 10 - 1 = 9$ da variável $B = (100,115,116,117,119)$, com amplitude $= 119 - 110 = 9$. A amplitude aumenta com o tamanho amostral: quanto mais observações, maior a probabilidade de encontrar valores extremos. No caso dos morcegos, a amplitude da variável do peso é $30 - 13 = 17$ g.

5.3.2 Soma de quadrados

Essa medida não é realmente uma medida de dispersão, mas deve ser mencionada porque forma a base para a variância e o desvio-padrão. Além disso, há algumas análises não paramétricas baseadas em permutação que se utilizam da soma de quadrados como estatística para análises. Calcula-se subtraindo a média amostral de cada medida individual e elevando a diferença ao quadrado. Elevar ao quadrado torna todos os valores positivos e há outras vantagens estatísticas. Depois de fazer isso para cada observação, somam-se esses desvios ao quadrado. Essa soma dos desvios da média ao quadrado é a soma de quadrados. Suponhamos que haja três observações $X = (3, 2, 1)$. A média é $3+2+1/3 = 2$. A soma de quadrados será $(3-2)^2 + (2-2)^2 + (1-2)^2 = 2$. Para os dados dos morcegos, a soma de quadrados da variável peso é 1771,43 g.

5.3.3 Variância da amostra

Para conseguir uma estimativa não tendenciosa da variância de uma população estatística com base numa amostra, divide-se a soma de quadrados por $n - 1$, sendo “n” o tamanho amostral. Então, a variância é

$$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}$$

A variância da variável peso dos morcegos é 13,95 g².

5.3.4 Desvio-padrão

A variância, apesar de ter propriedades estatísticas úteis e ser a base de muitos testes estatísticos, está em unidades ao quadrado. Um conjunto de comprimentos medidos em centímetros terá uma variância expressa em centímetros quadrados, o que é estranho. A raiz quadrada da variância é o desvio-padrão e fornece uma medida de dispersão que está nas unidades originais dos dados.

Além de ser mais facilmente entendido do que a variância como uma medida da variação dos dados, o desvio-padrão resume o quão próximas as observações estão da média de forma direta. O desvio-padrão do peso dos morcegos é 3,74 g. Então se alguém diz que o peso médio dos morcegos é 21,9 g com desvio-padrão de 3,74 g, sabe-se que aproximadamente dois terços dos morcegos do fragmento florestal provavelmente têm comprimento de corpo entre 25,64 e 18,16 g.

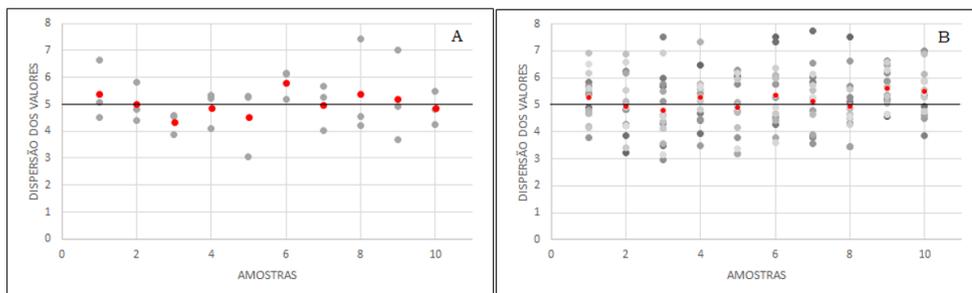
5.3.5 Coeficiente de variação

O coeficiente de variação é o desvio-padrão dividido pela média. Resume a quantidade de variação como uma percentagem ou proporção do total. É útil quando se compara a dispersão de valores de uma variável entre grupos com médias diferentes, ou entre diferentes variáveis quantitativas. O coeficiente de variação do peso de morcegos é 0,17 ou 17%.

5.3.6 Erro-padrão

Quando se toma uma amostra de medidas de um universo amostral e se calcula a média dessa amostra, se tem uma evidência do valor do parâmetro que descreve todo o universo amostral. A média da amostra não será exatamente igual à média de todo o universo amostral, nem igual a outra média calculada de outra amostra. Então, é desejável que se tenha uma ideia do quão próxima a média da amostra está da média “real”. Se o tamanho amostral é pequeno, a estimativa da média não será tão boa quanto uma estimativa baseada em um tamanho amostral maior. A seguir, são mostradas 10 amostras aleatórias de dados simulados com uma média “real” de 5.

Figura 13 – Estimativa de médias amostrais baseadas em uma população com média real = 5



Fonte: Elaborada pelo autor (2021).

A: Amostras com 3 objetos

B: Amostras com 17 objetos

Observando a Figura 13, pode se ver que, com um tamanho amostral de apenas 3 repetições da unidade amostral (gráfico A), algumas das médias amostrais (pontos vermelhos) não são muito próximas da média real. Com 17 observações por amostra, as médias das amostras são, em geral, mais próxi-

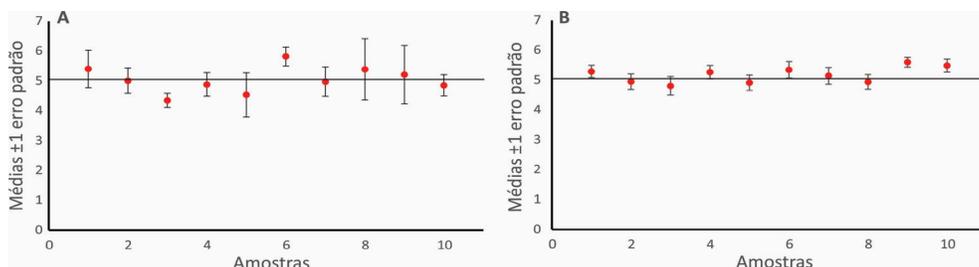
mas da média real (gráfico B). Quando se calcula a média de uma amostra, as pessoas têm de saber o quanto a média da amostra está próxima da média do universo amostral. Uma forma de se fazer isso é calculando o erro-padrão da média. Se forem tomadas muitas amostras aleatórias de um universo amostral, o erro-padrão da média é o desvio-padrão das diferentes médias das amostras. Aproximadamente dois terços (68,3%) das médias das amostras estão entre um erro-padrão da média populacional, 95,4% entre dois erros-padrões e quase todas médias (99,7%) entre três erros-padrões.

Quando se realiza um experimento, não há múltiplas amostras para estimar múltiplas médias e calcular o erro-padrão. Felizmente, pode-se estimar o erro-padrão da média usando o tamanho amostral e o desvio-padrão de uma única amostra. O erro-padrão da média é estimado pelo desvio-padrão das observações dividido pela raiz quadrada do tamanho amostral.

$$EP = \frac{s}{\sqrt{n}}$$

Onde EP = erro-padrão, s = desvio-padrão da amostra e n = tamanho amostral. Com tamanhos amostrais maiores, a média amostral se torna uma estimativa mais acurada da média populacional estatística, então o erro-padrão da média se torna menor. Note que isso é o que acontece nos gráficos da Figura 14. Quanto maior o tamanho amostral ($n = 17$), menor o erro-padrão da média.

Figura 14 – Médias e erros-padrões para amostras de uma população com média real = 5



Fonte: Elaborada pelo autor (2021).
A: Amostras com 3 objetos
B: Amostras com 17 objetos

Note que o erro-padrão é uma função da raiz quadrada do tamanho amostral. Então, para deixar o erro-padrão pela metade, é necessário quatro vezes mais unidades amostrais. Por exemplo, para diminuir o erro-padrão de uma amostra com três objetos pela metade, necessita-se ter 12 objetos. O erro-padrão da média para a variável peso dos morcegos é 0,33 g. Isso quer dizer que a média “real” do peso dessa população de *Sturnira lilium* no fragmento florestal amostrado está compreendida entre os valores $21,98 \pm 0,33$ g, ou entre 21,64 e 22,03 g. Esse valor do erro-padrão mostra que a média observada é uma boa estimativa da média populacional, pelo menos para esse experimento.

5.4 DESVIO-PADRÃO OU ERRO-PADRÃO?

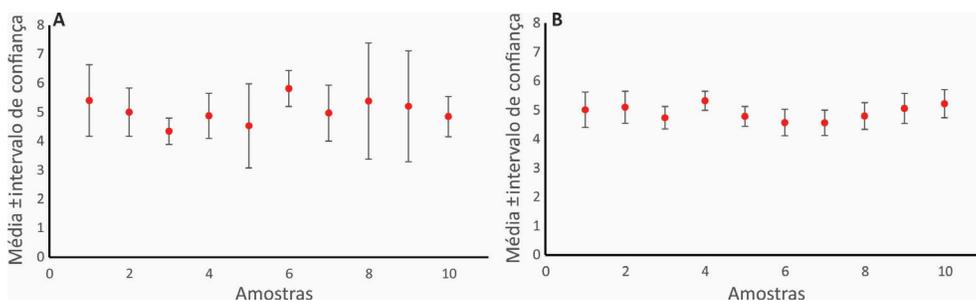
Usa-se o desvio-padrão e o coeficiente de variação para mostrar quanta variação há entre observações individuais. Por outro lado, usa-se o erro-padrão ou o intervalo de confiança para mostrar quanto sua estimativa da média “real” é boa. O único momento em que se reporta o desvio-padrão ou o coeficiente de variação é se houver interesse na quantidade de variação da amostra. Por exemplo, se alguém plantar um punhado de pés de milho com dois tipos de fertilizantes, seu interesse pode ser comparar as colheitas de milho com os dois fertilizantes. Então, se mostra a média das colheitas \pm o erro-padrão ou o intervalo de confiança. Se alguém quer selecionar artificialmente plantas de milho maiores para ter colheitas melhores, estará interessado em qual tipo de fertilizante tem a maior variação (tornando mais fácil escolher as plantas de milho que crescem mais), então se mostra o desvio-padrão ou o coeficiente de variação.

5.5 INTERVALOS E LIMITES DE CONFIANÇA

Intervalos de confiança e erros-padrões da média servem ao mesmo propósito: expressar a confiabilidade da estimativa da média. Limites de confiança indicam quanto sua estimativa da média populacional é precisa. Limites de confiança são os números na parte de cima e de baixo de um intervalo de confiança. Por exemplo, se a média é 7,4, com limites de confiança de 5,4 e 9,4, o intervalo de confiança é 5,4 a 9,4. A maioria das pessoas usam limites de confiança de 95%, mas se pode utilizar outros valores. Determinando 95% de limites de confiança significa que se forem tomadas repetidas amostras aleatórias de um

universo amostral e calculadas as médias para cada amostra, o intervalo de confiança para 95% das amostras incluirá a média da população. Os gráficos da Figura 15 mostram as médias e os intervalos de confiança para 10 amostras com tamanho amostral = 3 e 10 amostras com tamanho amostral = 17 objetos. Da mesma forma que no erro-padrão, para tamanhos amostrais menores, há maior desvio da média observada em relação à média real. Assim que se aumenta o tamanho amostral, essa variação diminui. O intervalo de confiança apresenta a amplitude completa que um valor de um parâmetro pode assumir. Também revela a precisão da estimativa do intervalo – quanto mais estreito, mais preciso. Entretanto, o intervalo de confiança não mostra sobre se o parâmetro está dentro dele. Lembre-se sempre de que há uma incerteza associada com a estimativa, seja 5%, 1% ou 0,1%. O melhor uso do intervalo de confiança envolve a comparação de intervalos de confiança entre estudos diferentes para revelar o parâmetro real, independente se o intervalo de confiança inclui o valor do parâmetro da hipótese nula, ou se a estatística em torno de qual o intervalo de confiança é construído é estatisticamente significativa (ZHANG *et al.*, 2008). O intervalo de confiança para a média dos dados de peso dos morcegos é 21,32 a 22,63 g. Os valores são bem parecidos com os do erro-padrão. Esse resultado informa que se pode ter 95% de confiança de que os valores da média “real” de peso dos morcegos se situam entre os limites de confiança.

Figura 15 – Médias e intervalos de confiança (95%) para amostras com 3 objetos



Fonte: Elaborada pelo autor (2021).

Há um mito de que quando duas médias têm barras de erros-padrões ou intervalos de confiança que não se sobrepõem, as médias são significativamente diferentes (no nível de $p < 0,05$). Isso não é inteiramente verdade (LANZANTE, 2005). Em alguns casos, dois conjuntos de números podem ter barras de erros-padrões ou intervalos de confiança que não se sobrepõem, mas, mesmo assim, não serem diferentes quando se calcula um teste-t, por exemplo. Portanto, para

não incorrer em erros, sugere-se não se interpretar diferenças significativas pela comparação visual entre barras de erros-padrões ou intervalos de confiança, mas que se use o valor do teste estatístico e o valor de p .

Quando se lê artigos científicos, às vezes, as “barras de erros” nos gráficos ou o número que vem depois do \pm das médias nas tabelas representam o erro-padrão da média, enquanto em outros artigos eles representam intervalos de confiança de 95%. Seja qual for a estatística que se decida usar, deve-se deixar claro que tipo de barras de erros os gráficos representam. Não há necessidade de reportar o desvio-padrão juntamente com o erro-padrão. Desde que se reporte um dos dois, juntamente com o tamanho amostral (n), qualquer pessoa pode calcular um ou outro com essas informações.

6 TESTES ESTATÍSTICOS

A partir de agora, conheceremos as análises estatísticas. Todas as análises descritas a seguir possuem tutoriais para serem realizadas no programa estatístico R. O leitor pode ler sobre o teste e realizá-lo com os mesmos dados apresentados. Além disso, a partir das orientações fornecidas, o leitor pode inserir seus próprios dados e realizar seus próprios testes estatísticos. Antes de estudarmos os testes, é preciso esclarecer que a abordagem proposta aqui é básica, pois esse é o objetivo deste livro, ou seja, ser um manual prático de consulta. Se a pessoa desenhar o experimento e coletar de forma adequada as informações de seu sistema de estudo, é quase certo que os testes a seguir poderão ser utilizados sem maiores preocupações. Entretanto, é sempre bom ficar atento, pesquisar e consultar colegas ou, até mesmo, um profissional da área da estatística caso haja receios. É bom lembrar que a inferência que se faz a partir dos resultados de um experimento pode fazer a diferença em ações na área da saúde, tecnologia, educação e meio ambiente. Por isso, deve-se ter o máximo de cuidado necessário para reduzir os riscos de aceitar uma hipótese falsa.

6.1 GRAUS DE LIBERDADE

Testes de significância estatística mostram o conceito de graus de liberdade. Em termos práticos, o número de graus de liberdade associado com uma estatística é igual ao número de seus componentes independentes, ou seja, o número total de componentes usados no cálculo menos o número de parâmetros que se tem para estimar dos dados antes de fazer a estatística (LEGENDRE; LEGENDRE, 2012). Os componentes são os objetos ou unidades amostrais do experimento. Por exemplo, o número de graus de liberdade associado com uma variância é o número de observações menos um (denota-se $v = n - 1$): n objetos são usados no cálculo, mas um grau de liberdade é perdido porque a média “real” é estimada dos dados amostrados. Em outras palavras, é o número de valores ou quantidades independentes que podem ser atribuídas para uma dis-

tribuição estatística. O importante é ter em mente que, para todos os testes que trabalharemos, com exceção do próximo, há graus de liberdade que devem ser reportados nos resultados.

6.2 VARIÁVEIS NOMINAIS

Os testes para variáveis nominais apresentados são baseados na distribuição de proporções de variáveis aleatórias discretas. Para todos os testes, o pressuposto é que as medidas sejam independentes, ou seja, que a partir do resultado de uma medida não seja possível nenhuma inferência sobre o valor de outra medida. Os testes também precisam que haja pelo menos cinco observações esperadas para cada categoria da variável aleatória discreta. Esses testes são menos exigentes quanto aos pressupostos em comparação com os testes para variáveis quantitativas e, portanto, exigem menos aprofundamento. Isso não significa que sejam menos úteis ou importantes. Muitas hipóteses científicas podem ser testadas com eles. Para algumas áreas da biologia, sua utilidade pode ser, inclusive, subestimada.

6.2.1 Teste exato do qui-quadrado

O teste exato do qui-quadrado é usado quando há uma variável nominal com dois ou mais valores. Comparam-se as contagens observadas em cada categoria com as contagens esperadas. A hipótese nula estatística é que o número de observações em cada categoria é igual ao esperado teoricamente e a hipótese alternativa é que os números observados são diferentes do esperado.

O teste exato do qui-quadrado calcula uma estatística que mede quanto os dados observados se desviam do esperado. Depois, usa-se a distribuição qui-quadrado de probabilidades para estimar a probabilidade de se obter o valor do teste estatístico sob a hipótese nula. O teste é calculado elevando a diferença entre o número observado (O) e esperado (E) ao quadrado e dividindo pelo esperado. Depois, somam-se todas as observações. Quanto maior o desvio da hipótese nula, maior a diferença entre o observado e esperado.

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

O valor do teste está relacionado com a diferença entre o observado e o esperado. Quanto maior é a diferença, maior fica o valor do teste.

6.2.1.1 Exemplo do teste exato do qui-quadrado

Suponhamos que foram coletados 102 indivíduos de três espécies de gramíneas da família Poaceae em um campo e, neste exemplo, não se sabe nenhuma informação prévia sobre as suas distribuições. Uma forma de pensar esse experimento é considerar que os indivíduos das três espécies se distribuem equitativamente no campo. Ou seja, o esperado é que não haja diferença nas distribuições. Então, uma expectativa teórica para a distribuição das proporções de indivíduos das espécies seria $1/3$, $1/3$ e $1/3$. A Tabela 7 mostra os dados.

Tabela 7 – Número de indivíduos de três espécies de gramíneas em um campo

	<i>Andropogon lateralis</i>	<i>Rynchospora flexuosa</i>	<i>Axonopus ciccus</i>
Contagens	65	24	13

Fonte: Elaborada pelo autor (2021).

Nesse caso, há apenas uma variável nominal (espécies de gramínea). Note que já podemos desconsiderar a hipótese de distribuição equitativa dos indivíduos das três espécies. Então, o teste nos dirá o quanto essa diferença se desvia significativamente do esperado $1/3$ para cada espécie. O valor do teste qui-quadrado foi 44,18 com dois graus de liberdade, e o valor de $p = 2,554 \times 10^{-10}$. Essa forma matemática de apresentar o valor de p é para quando os valores são muito pequenos. Nesse caso, há 10 zeros antes do número 2,554 (0,0000000002554). Pode-se apresentar o resultado como $p < 0,0001$. A hipótese nula do teste é que as proporções são igualmente distribuídas. Já tínhamos visto que isso não era verdade. Mas o teste de hipótese nos deu uma representação do quanto o observado se desvia do esperado. Esse exemplo é bem básico e a maioria das pessoas concordará que não é preciso realizar um teste para dizer que as espécies são diferentes na distribuição de seus indivíduos, pelo menos para esse campo.

Outro objetivo, menos óbvio e mais funcional, que pode ser contemplado com esse tipo de teste é saber se há diferença na proporção das três espécies no campo do estudo em comparação com as proporções esperadas. Suponhamos que as distribuições dessas espécies já tenham sido estudadas em vários campos diferentes. Os resultados apontam para uma determinada distribuição

padrão para cada uma das espécies de 3:2:1 (leia-se 3 para 2 para 1). Ou seja, para cada três indivíduos de *Andropogon lateralis*, temos dois indivíduos de *Rynchospora flexuosa* e um indivíduo de *Axonopus ciccus*. Então, $3+2+1 = 6$ e as proporções para entrar na análise são $3/6$, $2/6$ e $1/6$.

O teste exato do qui-quadrado foi 7,73 com dois graus de liberdade, e o valor de $p = 0,021$. A hipótese nula do teste é que as proporções observadas são iguais às proporções esperadas. Note que, neste novo experimento, as proporções esperadas mudaram. Então, a hipótese de referência do estudo é que se espera uma diferença nas proporções. Neste caso, a hipótese de referência é igual à hipótese nula estatística. Com o resultado do teste, rejeita-se essa hipótese. Isso quer dizer que há uma diferença nas proporções de indivíduos dessas três espécies neste campo em comparação com o esperado. Para saber quais espécies são diferentes entre si, pode-se fazer testes separados. Esse tipo de análise é simples, mas pode ser utilizado para perguntas interessantes. Por exemplo, suponha que uma dessas espécies seja uma exótica invasora. O objetivo poderia ser observar se o número aumenta em relação a outro campo. Outro objetivo, caso o número aumente em relação ao mesmo campo ao longo do tempo. Um terceiro objetivo seria comparar com a distribuição nativa dessa espécie. Nesse caso, teríamos um experimento de monitoramento de uma espécie exótica invasora.

6.2.2 Teste qui-quadrado de independência

O teste qui-quadrado de independência é utilizado quando se tem duas variáveis aleatórias nominais, cada uma com dois ou mais valores possíveis. Os dados são usados para calcular um teste estatístico que mede o quão distante os dados observados estão dos valores esperados. Depois, se usa uma relação matemática para estimar a probabilidade de obter o valor do teste estatístico. No caso do qui-quadrado, a distribuição é qui-quadrado. Quanto maior o desvio da hipótese nula, maior será a diferença entre o observado e o esperado. O cálculo do teste é igual ao anterior, só que não há uma expectativa teórica, mas, sim, valores esperados calculados diretamente dos dados. A forma de apresentação dos dados para o teste qui-quadrado é a tabela de contingência. O teste exato de Fisher é mais acurado do que o teste de independência do qui-quadrado quando os valores esperados são baixos. Então, em geral, recomenda-se o teste do qui-quadrado se o tamanho amostral total é maior do que 1000 objetos, pois é um tamanho amostral que está na extensão em que um teste exato e o teste qui-quadrado fornecerão valores de p praticamente iguais (MCDONALD, 2014).

A hipótese nula do teste é que as proporções relativas de uma variável são independentes daquelas da segunda variável. Em outras palavras, as proporções em uma variável são iguais para diferentes valores da outra variável.

Mesmo que se tenha um tamanho amostral adequado, o teste qui-quadrado de independência ainda pode fornecer valores de p muito baixos, principalmente quando os valores esperados são pequenos. Uma solução para isso é utilizar a correção de Yates, também chamada de correção de continuidade, a qual aplica o teste qui-quadrado de independência depois de subtrair 0,5 de cada valor observado que é maior do que o valor esperado e adicionar 0,5 para cada valor observado que é menor do que o valor esperado. Essa correção se aplica somente para tabelas de contingência 2×2 .

6.2.2.1 Exemplo do teste qui-quadrado de independência

A Tabela 8 mostra o resultado das distribuições de aranhas arbustivas com as teias fixadas a folhas ou ramos em 16 fragmentos florestais de Mata Atlântica do sul do Brasil. Há uma pequena maior proporção de adultos que utilizam folhas (4,61%) do que ramos (3,41%). Há $1344 + 1086 + 62 + 37 = 2529$ observações e, portanto, o tamanho amostral é maior do que 1000 e utiliza-se o teste qui-quadrado. A hipótese nula é que a proporção de aranhas adultas que fixam teias em folhas é igual à proporção de aranhas adultas que fixam teias em ramos. Em outras palavras, não há preferência por folhas ou ramos por parte de aranhas adultas.

Tabela 8 – Contagens de pontos de contato de teias de aranhas jovens em folhas ou ramos em fragmentos florestais

	Jovens	Adultos	Total
Folhas	1344	62	1406
Ramos	1086	37	1123
Total	2430	99	2529

Fonte: Elaborada pelo autor (2021).

Calculam-se as frequências esperadas a partir das observadas. Das 2529 aranhas, $62 + 1344 = 1406$ escolheram folhas para fixar a teia. A hipótese nula é, portanto, que $62/1406 = 4,41\%$ das aranhas adultas escolhem folhas. A pergunta, agora, é se a diferença entre o observado de 4,61% aranhas adultas usando folhas e o esperado de 4,41% é grande o suficiente para dizermos que há uma preferência por folhas. O raciocínio é igual aos jovens usando folhas ou ramos.

O resultado do teste qui-quadrado foi 2,06. Os graus de liberdade em um teste de independência são iguais a $(\text{número de linhas}) - 1 \times (\text{número de colunas}) - 1$. Então, para uma tabela 2×2 , há $(2 - 1) \times (2 - 1) = 1$ grau de liberdade. Para um valor de teste de qui-quadrado de 2,06 com 1 grau de liberdade, o valor de p é 0,15. Aceitamos a hipótese nula e concluímos que aranhas adultas usam folhas e ramos com as mesmas proporções nos fragmentos florestais estudados.

Geralmente, a apresentação da tabela de contingência com os valores das proporções (como mostrado acima) deve bastar para apresentar os resultados. Também se pode montar um gráfico de barras com os valores de uma variável no eixo X e as proporções da outra variável no eixo Y. No caso do exemplo das aranhas, basta indicar qual a percentagem de aranhas adultas que utilizaram folhas ou ramos. O valor do teste, os graus de liberdade e o valor de p devem ser reportados.

6.2.3 Teste exato de Fisher

O teste exato de Fisher é usado quando há duas variáveis nominais e se quer saber se as proporções de uma variável são diferentes dependendo dos valores de outra variável. Esse teste deve ser escolhido no lugar do qui-quadrado quando o tamanho amostral for pequeno e/ou um dos números esperados for menor que 5. A hipótese nula do teste é que as proporções de uma variável são independentes da outra variável. Em outras palavras, as proporções em uma variável são iguais para diferentes valores da segunda variável.

Em contraste com o teste qui-quadrado de independência, o teste exato de Fisher não se utiliza de uma distribuição particular para calcular um valor de teste. A associação entre duas variáveis em uma tabela de contingência pode ser inferida a partir de vários métodos. O teste exato de Fisher apresentado aqui se utiliza da razão de possibilidades (*odds ratio*). Dados dois eventos, A e B, a razão de possibilidades é definida como a razão das chances de ocorrer o evento A na presença do evento B e a possibilidade de A na ausência de B. Podemos pensar de forma equivalente, que é a razão das chances de B ocorrer em presença de A ou a possibilidade de B ocorrer na ausência de A. Dois eventos – ou categorias – são independentes se, e somente se, a razão de possibilidade é 1. Se a razão é maior do que 1, os eventos são associados positivamente e, se a razão é menor do que 1, os eventos são negativamente associados.

Para uma tabela de contingência de tamanho 2×2 , as probabilidades são:

	B = 1	B = 0
A = 1	p_{11}	p_{10}
A = 0	p_{01}	p_{00}

E a razão das possibilidades é:

$$RP = \frac{p_{11} p_{00}}{p_{10} p_{01}}$$

Para o teste exato de Fisher, podemos calcular valores de limites de confiança. Como sabemos que a razão de possibilidade da hipótese nula de ausência de associação entre as variáveis é 1, se esse valor estiver dentro do intervalo de confiança, aceitamos a hipótese nula. Se o valor 1 estiver fora do intervalo, rejeitamos a hipótese nula. Há diferentes definições quanto ao significado de um intervalo de confiança. Para a abordagem frequentista, um intervalo de confiança de 95% significa que se realizarmos várias amostragens do mesmo universo amostral com o mesmo tamanho amostral e calcularmos os intervalos de confiança, podemos esperar que 95% dos intervalos contenham a proporção real do universo amostral e apenas 5% dos intervalos não a contenham (SOKAL; ROHLF, 2009). Ou seja, o intervalo de confiança conterá a proporção real em 95 de 100 vezes.

6.2.3.1 Exemplo do teste exato de Fisher

A Tabela 9 mostra o resultado de uma pesquisa realizada por estudantes de graduação de um curso de Ciências Biológicas. O objetivo do estudo foi testar se a prática de exercícios físicos está associada ao sexo. Para tanto, 50 pessoas foram entrevistadas nos corredores da universidade.

Tabela 9 – Contagens de homens e mulheres que praticam ou não exercício físico

	Sim	Não	Total
Homem	15	8	23
Mulher	9	18	29
Total	24	26	50

Fonte: Elaborada pelo autor (2021).

O teste exato de Fisher fornece o valor da probabilidade (valor de p) de que haja uma associação entre as duas variáveis nominais. Do total de 24 pessoas que responderam “Sim” para a prática de exercícios físicos, 9 eram mulheres e 15 eram homens. O valor de p para o teste exato de Fisher é 0,046. Note que esse é um resultado em que o valor de p ficou muito próximo do valor de corte de $p = 0,05$. Ainda assim, como todos os resultados são comparados com o nível de significância estabelecido, devemos aceitar a hipótese alternativa de que há uma associação entre o sexo e a prática de exercícios físicos. O intervalo de 95% de confiança foi 1,007669 a 14,320910 e a razão de possibilidades foi 3,644563. Note que o intervalo de confiança calculado *não* englobou a proporção 1 por muito pouco. Assim como o valor de p , essa estatística ficou no limite. A razão de possibilidades mostra que a chance de homens fazerem exercício é em torno de 3,6 vezes maior do que de mulheres realizá-los.

Note que o pressuposto do teste é que as observações são independentes. No exemplo, isso significa que a resposta de uma pessoa não pode influenciar a resposta de outra pessoa. Se os estudantes aplicassem o questionário no curso de Educação Física, todas as pessoas, independente do sexo, diriam que realizam exercícios. Ou seja, as respostas não seriam independentes e o resultado do teste seria tendencioso.

6.2.4 Teste de McNemar

O teste de McNemar é utilizado quando se tem uma tabela de contingência com duas variáveis nominais. Uma das variáveis é resultado de uma amostragem pareada, ou seja, quando uma observação da primeira amostra é pareada com uma observação da segunda. Esse tipo de amostragem é utilizada quando se quer saber se uma intervenção realizada sobre um conjunto de objetos surtiu efeito (foi significativa). Então, uma variável nominal é o *antes* e *depois* e a segunda variável nominal obrigatoriamente deve ser *sim* e *não*.

A estatística utilizada é o qui-quadrado de McNemar, que utiliza as proporções das respostas que mudaram de um estado para o outro estado. Além da independência, outro pressuposto do teste é o pareamento das amostras, ou seja, que os dados venham do mesmo conjunto de objetos.

6.2.4.1 Exemplo de teste de McNemar

Suponhamos que um adestrador de cães inventou um novo método para alterar o comportamento incorreto dos cães de pular nas pessoas quando elas entram em seus territórios e quer colocá-lo a um teste preliminar. Ele seleciona um conjunto de cães e pergunta aos donos se os animais possuem o comportamento correto ou incorreto. Essa é a primeira amostra. Então, ele aplica o novo método durante o tempo que pensa ser necessário para alterar o comportamento. Depois, devolve esses cães para seus donos. Passados alguns dias, o adestrador pergunta novamente aos donos se os animais têm o comportamento correto ou incorreto. Essa é a segunda amostra com os mesmos objetos (pessoas, no caso). A Tabela 10 mostra o resultado do questionário.

Tabela 10 – Resultado da intervenção no comportamento de cães

Antes	Depois	
	Correto	Incorreto
Correto	10	1
Incorreto	20	9

Fonte: Elaborada pelo autor (2021).

Como mostra a Tabela 11, temos quatro opções para os valores de antes e depois da aplicação do método.

Tabela 11 – Valores pareados para os dados da intervenção

Antes	Depois	
Correto	Correto	10
Correto	Incorreto	1
Incorreto	Correto	20
Incorreto	Incorreto	9

Fonte: Elaborada pelo autor (2021).

Note que, antes, havia 15 cães que se comportavam de forma correta e 29 cães que se comportavam de forma incorreta. Depois da aplicação do método, um cão alterou o comportamento correto para o incorreto e 20 cães mudaram o comportamento de incorreto para correto. O resultado do teste do qui-quadrado de McNemar = 17,19 e o valor de $p = 0,000034$, o que indica que

o método aplicado funcionou na mudança de comportamento dos cães. Assim como para o teste qui-quadrado de independência, para o teste de McNemar, se sugere aplicar uma correção, chamada correção de continuidade, quando há valores menores do que 5. No exemplo, há uma observação com o valor 1. O valor do teste de McNemar corrigido foi 15,43 com um valor de $p = 0,000086$.

6.3 VARIÁVEIS QUANTITATIVAS

Os testes estatísticos para variáveis aleatórias quantitativas apresentados aqui são chamados univariados. São aqueles em que há apenas uma variável dependente ou resposta, e uma ou mais variáveis independentes ou explicativas que podem afetar o comportamento da variável resposta.

Antes de escolher o teste estatístico, também é importante que o pesquisador considere a natureza da sua hipótese de estudo. A intenção da hipótese é testar por diferenças em médias entre grupos, ou é testar por associações entre variáveis? Testar diferenças entre médias significa fazer inferências sobre se os valores de uma variável aleatória quantitativa são diferentes entre dois ou mais grupos. Testar associações entre duas ou mais variáveis significa que se está interessado em saber o quanto uma variável Y varia em função de uma (ou mais) variável X .

6.3.1 Pressupostos estatísticos

O que é pressuposto estatístico? Pressuposto de um teste tem a ver com algumas características numéricas que os dados devem apresentar. As análises de dados clássicas (ou paramétricas) exigem que os dados coletados se comportem de algumas formas particulares. Porém, para outras abordagens, isso não é inteiramente necessário. Nas próximas seções, aprenderemos sobre alguns pressupostos básicos aos quais devemos ficar atentos. Nas seções destinadas aos testes, sempre que for necessário, retomaremos o assunto para aprofundar as necessidades de cada teste com relação aos pressupostos básicos e a outros pressupostos específicos.

6.3.1.1 *Independência das observações*

Na parte dos testes para variáveis nominais, já citamos esse pressuposto. A independência de observações é quando o resultado da medida em um objeto não é influenciado pela medida em outro objeto, tanto no âmbito espacial, quanto no temporal. No caso espacial, o oposto da independência é a correlação espacial. Isso produz relações entre valores observados em locais vizinhos no espaço e, portanto, causa uma falta de independência espacial dos valores observados para as variáveis. Em muitos casos, as observações que são mais próximas tendem a ter valores que são mais similares do que observações mais distantes, resultando em dependência espacial positiva, ou correlação espacial positiva. Então, a falta de independência espacial entre observações cria um problema na utilização de testes de significância que assumem independência de observações. Quando isso acontece, se diz que há pseudorrepetição (HURLBERT, 1984). Uma das abordagens para lidar com a falta de independência utilizada durante a fase de desenho amostral é a seleção aleatória dos objetos ou a alocação aleatória de tratamentos aos objetos (OEHLERT, 2010).

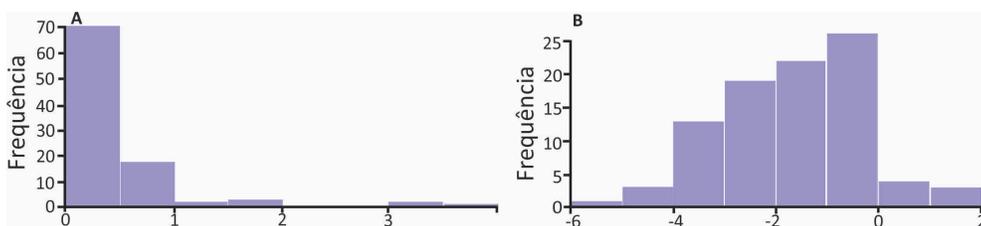
Por exemplo, suponhamos um experimento para examinar os efeitos das condições ambientais de um sítio no tamanho da distribuição de uma espécie de inseto galhador (SCHEINER; GUREVITCH, 2001). Galhas se formam em caules de arbustos com cada galha contendo várias câmaras com uma larva em cada uma. As galhas são coletadas e as larvas são medidas de uma forma hierárquica: 3 larvas são medidas em cada galha; 5 galhas são coletadas de cada galho de cada arbusto; 20 arbustos são amostrados em cada sítio; 4 sítios são amostrados. Qual seria o nível mínimo da repetição verdadeira? Em outras palavras, qual é a unidade amostral independente? A resposta é “depende do objetivo”. Para cada larva dentro da galha, o crescimento de uma afeta o crescimento da outra, ou seja, cada larva dentro de uma galha não é independente. Isso é um caso de pseudorrepetição. Entretanto, para as galhas crescendo em diferentes galhos, apesar das observações não serem independentes (estão no mesmo arbusto), isso não é um caso de pseudorrepetição. Isso significa, de forma prática, que a confiança nos resultados de testes estatísticos realizados em objetos que não possuem independência não é a mesma. Diretamente falando, pode-se ser enganado pelos resultados do teste, o que não é desejável de jeito nenhum. Ao contrário de outros pressupostos dos testes estatísticos, não é fácil olhar os dados e ver se são independentes ou não. É necessário entender o sistema de estudo e desenhar cuidadosamente o experimento para que as observações sejam independentes. Alternativamente,

pode-se realizar um experimento piloto. Há estatísticas que medem a quantidade de correlação entre as medidas de uma variável, que não serão abordados aqui. Além da dependência espacial, quando dados são coletados ao longo do tempo, pode haver dependência temporal. Há testes estatísticos desenvolvidos especialmente para séries temporais, que também não serão abordados.

6.3.1.2 Normalidade

Já estudamos a distribuição normal. Também vimos que um histograma é a ferramenta gráfica indicada para analisar a distribuição de probabilidades de uma variável aleatória quantitativa. Estudaremos um pouco mais o que significa esse pressuposto para um experimento. A Figura 16 mostra dois histogramas para a variável aleatória quantidade de fósforo na água de um reservatório de uma barragem no sul do Brasil. No gráfico A, notamos que a distribuição está enviesada à esquerda, ou seja, não é normal. O gráfico B é o resultado da transformação em logaritmos dos dados de fósforo e pode-se ver que a forma da distribuição está mais parecida com a forma de sino da distribuição normal.

Figura 16 – Histogramas de distribuição de frequências de medidas de fósforo



Fonte: Elaborada pelo autor (2021).

A: Dados sem transformação

B: Dados com transformação logarítmica

Se a variável quantitativa não é normalmente distribuída, pode-se estar aumentando a probabilidade de um resultado falso positivo se os dados forem analisados com testes que assumem normalidade. Se você quer analisar seus dados com esses testes, assim que os dados forem coletados, deve olhar o histograma de frequências para ver se parece com uma curva normal. O melhor é

realizar um experimento piloto, que consiste em coletar alguns dados, checar a normalidade e outras características dos dados e proceder com o experimento para valer. Fazer um piloto pode ser pouco complicado em experimentos de laboratório, um pouco mais complicado em experimentos de campo e muito complicado em experimentos observacionais. Portanto, tudo dependerá do tipo de experimento e dos objetivos do estudo.

Se os dados parecem não normais, mesmo que se faça uma transformação, provavelmente ainda é adequado analisá-los usando um teste clássico. Eles não são assim tão sensíveis à não normalidade (não há um consenso de que a não normalidade é necessária para invalidar um experimento). Entretanto, há testes alternativos. Por exemplo, um teste de permutação foi utilizado para comparar os métodos pós-operatórios no início do livro.

6.3.1.3 Homocedasticidade

Homocedasticidade significa que as variâncias são iguais. Quando comparamos médias entre grupos, as variâncias de cada grupo devem ser iguais. Quando estudamos a associação entre variáveis em um teste de regressão, esse pressuposto é um pouco mais complicado de explicar. Um teste de regressão é uma análise que gera um modelo de comportamento da variável resposta em função da variável explicativa. A igualdade de variâncias, neste teste, tem a ver com a variância dos valores observados e a variância dos valores preditos para o modelo. Na seção sobre regressão, retornaremos a esse conceito. Há maneiras gráficas e analíticas de testar a homocedasticidade. O teste analítico mais popular é o teste de Bartlett, que pode ser usado tanto para análises baseadas em comparações entre médias, quanto em associações entre variáveis.

6.3.2 Teste-t para duas amostras

Há vários testes estatísticos que usam a distribuição de probabilidade t e podem ser chamados teste- t . Um dos mais comuns é o teste- t de Student. Student é o pseudônimo de William Gosset, funcionário da cervejaria Guinness, empresa que não deixava seus funcionários publicarem artigos, mas abriu uma exceção para Gosset e não queria que os colegas soubessem.

Utiliza-se o teste- t de Student quando se tem uma variável quantitativa e uma variável nominal com dois grupos e o objetivo é comparar as médias desses grupos. Então, a variável nominal deverá possuir somente dois valores, tal

como macho e fêmea, com e sem tratamento, rural e urbano, etc. A hipótese nula estatística desse teste é que as médias da variável quantitativa são iguais para os dois grupos.

O teste estatístico é calculado como (QUINN; KEOUGH, 2002):

$$t_s = \frac{\bar{x}_1 + \bar{x}_2}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Onde t_s = valor do teste- t , no denominador temos as duas médias, e s_1^2 e s_2^2 são as duas variâncias e n_1 e n_2 são os tamanhos amostrais dos dois grupos. Isso quer dizer que t_s é maior quanto maior for a diferença entre as médias, quanto menor a variância das amostras e quanto maior for o tamanho amostral. Calcula-se a probabilidade de se obter o valor observado de t_s sob a hipótese nula usando a distribuição de probabilidades t . A forma da distribuição t e, portanto, a probabilidade de se obter um valor particular de t_s , depende do número de graus de liberdade. O número de graus de liberdade de um teste- t é o número total de observações nos grupos menos 2, ou $n_1 + n_2 - 2$.

O teste- t não é sensível a desvios da normalidade se as distribuições dos dois grupos forem iguais. Isso quer dizer que, mesmo sem cumprir o pressuposto de normalidade, ainda se pode confiar na probabilidade de 5% de rejeitar erroneamente a hipótese nula, ou incorrer em um erro Tipo I. Se os dados são muito não normais e se há dois grupos com distribuições diferentes (por exemplo, um com distribuição normal; outro com distribuição binomial) e, além disso, se tem tamanhos amostrais pequenos (menos que 50), então o teste- t para duas amostras fornecerá resultados com consideravelmente mais que 5% de falsos positivos, o que é inaceitável. Uma transformação de dados não ajudará nesse caso e nem mesmo um teste não paramétrico como o de Mann-Whitney.

Se o desenho é balanceado (tamanhos amostrais iguais nos dois grupos), o teste não é muito sensível a desvios da homocedasticidade, a menos que o tamanho amostral seja muito pequeno (menos que 10). Com um desenho desbalanceado (tamanhos amostrais diferentes nos dois grupos), um desvio da homocedasticidade é um problema. Se os dois grupos possuem desvios-padrões que são substancialmente diferentes (tal como um desvio padrão é três vezes maior que o outro) e os tamanhos amostrais das amostras forem pequenos (menor que 10) ou diferentes, deve-se usar o teste- t de Welch.

Para o teste-*t*, podemos calcular valores de limites de confiança. Nesse caso, um intervalo de confiança descreve a amplitude de valores que pode conter a diferença real entre as médias dos dois grupos. Para um teste-*t*, a hipótese nula é que a diferença entre as médias é igual a zero, ou seja, não há diferença. A hipótese alternativa é que a diferença real entre as médias é diferente de zero.

6.3.2.1 Exemplo de teste-*t* de Student

Analisaremos o resultado de um exercício. No outono de 2004, foram amostradas as alturas de duas turmas de estudantes da disciplina de Análise de Dados em Biologia de uma universidade. A turma A apresentou uma média de altura de 169,58 cm e desvio-padrão 12,59 cm, e a turma B apresentou média de altura de 163,90 cm com desvio-padrão de 9,07 cm. A diferença entre as médias, portanto, foi de 5,68 cm. As duas médias são significativamente diferentes? Os dados são apresentados na Tabela 12.

Tabela 12 – Alturas de estudantes de duas turmas da disciplina de Análise de Dados em Biologia

A	B
175,26	172,72
177,80	157,48
167,64	170,18
160,02	172,72
172,72	175,26
177,80	170,18
175,26	154,94
170,18	149,86
157,48	157,48
160,02	154,94
193,04	175,26
149,86	167,64
157,48	157,48
157,48	157,48
190,50	154,94
157,48	177,80
182,88	160,02

Fonte: Elaborada pelo autor (2021).

Há uma variável quantitativa (altura) e uma variável nominal (turma) dividida em dois grupos. A hipótese nula é que as alturas médias são iguais nas duas classes. O resultado do teste- t foi $t_s = 1,508$, graus de liberdade = 33, $p = 0,14$, que não rejeita a hipótese nula e, portanto, as duas médias são estatisticamente iguais. O intervalo de confiança foi -1,989 a 13,345, o que compreende o zero e, portanto, confirma a aceitação da hipótese nula.

Por se tratar da comparação de dois números, raramente se colocam os resultados de um teste- t em um gráfico para publicação. Para uma apresentação, pode-se montar um gráfico de barras com as médias e os desvios-padrões ou erros-padrões, como ilustrado para a análise de variância de uma via no próximo tópico. Por outro lado, deve-se apresentar as médias e os valores de alguma medida de dispersão, dependendo do objetivo do estudo, juntamente com o valor do teste, os graus de liberdade e o valor de p .

6.3.3 ANOVA de uma via

A análise de variância ou ANOVA (**AN**alysis **Of** **VA**riance em inglês) é a técnica mais utilizada para comparar as médias de grupos de uma variável aleatória quantitativa. Há vários tipos de desenhos amostrais que podem ser analisados com diferentes tipos de ANOVAs. No presente livro, estudaremos a ANOVA de uma via, ANOVA de duas vias e a ANOVA aninhada.

Em uma ANOVA de uma via (também conhecida como de um fator ou de fator único), há uma variável quantitativa e uma variável qualitativa nominal com dois ou mais grupos. A hipótese nula estatística é que as médias da variável quantitativa são iguais para os diferentes grupos. Quando há duas médias, é bem tranquilo declarar a hipótese de estudo, ou algo sobre qual média se pensa ser maior. Mas, com mais de duas médias, as coisas ficam um pouco mais complicadas. Por exemplo, se alguém quiser comparar cinco médias, o ideal é desenhar uma hipótese a respeito de qual grupo teria a maior média, depois, a partir disso, determinar quais grupos teriam médias cada vez menores. Fazendo isso, se pensará na lógica de resposta esperada do sistema de estudo.

A ideia básica da análise de variância é calcular a média das observações dentro de cada grupo e, então, comparar a variância dentro dos grupos com a variância média entre cada grupo. Sob a hipótese nula de que todas as observações nos diferentes grupos terão a mesma média, a variância balanceada entre grupos será a mesma do que a variância dentro do grupo. Na medida que as médias ficam mais distantes entre si, a distância entre as variâncias das médias

de cada grupo aumenta. O teste estatístico que infere o tamanho do efeito das diferenças entre as médias é o teste F , que se baseia na razão da variância entre as médias (*entre grupos*) e a média das variâncias *dentro* dos grupos. Quanto maior a variância entre os grupos, maior o valor de F , ou seja, maior o tamanho do efeito da diferença entre as médias. Essa estatística possui uma distribuição de probabilidades F conhecida sob a hipótese nula e, então, a probabilidade de obter o valor de F observado pode ser calculada.

A forma da distribuição F depende de dois graus de liberdade: os graus de liberdade do numerador (variância *entre* grupos) e os graus de liberdade do denominador (variância *dentro* de grupos). Os graus de liberdade entre os grupos são o número de grupos menos um. Os graus de liberdade dentro dos grupos correspondem ao total do número de observações (n) menos o número de grupos. Então se há n observações em a grupos, os graus de liberdade do numerador correspondem a $a - 1$ e do denominador a $n - a$.

A ANOVA pressupõe que as observações dentro de cada grupo sejam normalmente distribuídas. O teste não é particularmente sensível a desvios dessa pressuposição (RASCH; GUIARD, 2004). Se uma ANOVA de uma via for aplicada para dados que não desviem “muito” da normalidade, a chance de obter um valor de p menor que 0,05, se a hipótese nula for verdadeira, é ainda muito próximo de 0,05. É melhor que os dados estejam distribuídos de forma aproximadamente normal, então, após coletar os dados é aconselhável fazer um histograma e aplicar o teste de Shapiro-Wilk. Por outro lado, se tiver poucos objetos, um histograma não será útil.

É possível transformar os dados, caso haja muito desvio da normalidade. Se nenhuma das transformações funcionar pode ser utilizado o teste de Kruskal-Wallis. Mas esse teste não paramétrico pressupõe que os diferentes grupos possuem as mesmas formas de distribuição (não normais, mas iguais). Este teste também *não* testa a mesma hipótese que uma ANOVA de uma via, pois não usa as médias. Há outros testes alternativos que se utilizam, por exemplo, da soma de quadrados como estatística e que geram um valor de p baseado em permutações. Esses testes não pressupõem que os dados sejam normalmente distribuídos.

Os desvios-padrões nos diferentes grupos devem ser observados para verificar se há grande diferença entre eles. Resultados iniciais sugeriram que desvios da homocedasticidade não eram um problema com desenhos balanceados (GLASS *et al.*, 1972), mas resultados posteriores encontraram que grandes desvios podem aumentar a taxa de falsos positivos, mesmo quando os tamanhos amostrais são iguais (HARWELL *et al.*, 1992). Em geral, dados com diferentes desvios mostrarão mais falsos positivos (rejeitar a hipótese nula quando ela é

verdadeira) para um desenho balanceado somente se um desvio-padrão for ao menos três vezes o tamanho do outro desvio-padrão e o tamanho amostral em cada grupo for menor do que 10.

Desvios da homocedasticidade são problemas maiores quando se tem um desenho desbalanceado (tamanhos amostrais diferentes nos grupos). Se os grupos com menores tamanhos amostrais também possuírem maiores desvios-padrões, haverá muitos falsos positivos. A diferença nos desvios-padrões não precisa ser grande. Um pequeno grupo pode ter um desvio-padrão 50% maior e sua taxa de falsos positivos pode ser acima de 10%, ao invés do aceitável 5% (0,05). Se os grupos com maiores tamanhos amostrais possuem os maiores desvios-padrões, o erro está na direção inversa, ou seja, haverá pouquíssimos falsos positivos, o que pode parecer uma coisa boa, exceto que isso também significa que se perde poder (terá muitos falsos negativos, se há uma diferença nas médias). Falsos negativos significam que se aceita a hipótese nula com mais frequência quando ela não é verdadeira.

A indicação é que se deve montar experimentos balanceados para todos os grupos. Com isso, pode-se seguramente utilizar uma ANOVA, a não ser que os tamanhos amostrais por grupo sejam inferiores a 10 e os desvios-padrões variem em 3 ou mais vezes. Nesse caso, há muitos testes alternativos para escolher (LIX *et al.*, 1996). O teste indicado para substituir uma ANOVA é a ANOVA de Welch.

Para uma ANOVA, podemos calcular intervalos de confiança em um teste de contrastes para cada par de grupos, caso o resultado do teste global seja significativo. A lógica para rejeitar a hipótese nula de que dois grupos possuem médias iguais é igual à do teste-*t*. Ou seja, se o intervalo de 95% de confiança contém o zero, isso significa que as médias são iguais. Os intervalos de confiança das médias também são um indicativo da amplitude de variação dentro da qual podemos encontrar a média “real” com 95% de confiança.

6.3.3.1 Exemplo de ANOVA de uma via

Baldissera *et al.* (2008a) coletaram aranhas arbustivas em 12 manchas florestais ($n = 12$) distribuídas em quatro grupos ($a = 4$): floresta nativa, floresta plantada de *Araucaria angustifolia*, floresta plantada de *Pinus* e floresta plantada de *Eucalyptus*. Entre os objetivos estava comparar a média de abundância (número de indivíduos) de aranhas entre os grupos. A hipótese era de que a floresta nativa comportaria maior abundância, pois haveria maior diversidade estrutural da vegetação nessa área e, portanto, mais estruturas para que mais aranhas pudessem utilizar. Uma maneira convencional de apresentar o resultado do teste é com a Tabela 13.

Tabela 13 – Resultado da ANOVA de uma via

Efeito	Soma de quadrados	Graus de liberdade	Quadrados médios	F	p
Entre grupos	18406,7	3	6135,56	4,968	0,031
Dentro de grupos	9879,3	8	1234,92		
Total	28286	11			

Fonte: Baldissera *et al.* (2008a).

Como há quatro grupos, o número de graus de liberdade entre grupos é 3 (4-1) e, como há 12 objetos, o número de graus de liberdade dentro de grupos é 8 (12-4). O valor de F é significativo (valor de $p = 0,031$). Significativamente igual ou diferente não quer dizer nada, biologicamente falando. Por outro lado, indicar que em um lugar há uma média maior do que em outro faz mais sentido. Para isso, precisamos realizar outro teste estatístico, pois a ANOVA apenas mostrou que há diferenças entre as 4 médias, mas não mostrou qual média é diferente de qual.

6.3.3.1.1 Teste de Tukey-Kramer

Para saber quais médias são diferentes entre si, é necessário realizar um teste de contrastes (também chamado teste *post-hoc*). O importante na escolha de um teste de contrastes é que ele não seja afetado pelas comparações múltiplas realizadas com os mesmos dados. A natureza probabilística de variáveis aleatórias determina que, se realizarmos muitos testes com o mesmo conjunto de dados, a probabilidade de encontrar algum tipo de efeito significativo aumentará. Isso significa que estamos artificialmente aumentando a chance de encontrar um efeito quando ele não existe, o erro Tipo I. É por isso que não podemos simplesmente realizar vários testes- t para saber quais médias são diferentes em substituição a uma ANOVA.

Quando há a grupos, há $(a^2 - a)/2$ comparações possíveis. No exemplo, há $(4^2 - 4)/2 = 6$ comparações possíveis. Se há 5 grupos, há 10 comparações. Há vários testes diferentes para comparações dos grupos par a par depois de uma ANOVA de uma via e cada um deles tem suas vantagens e desvantagens. As diferenças entre seus resultados são sutis e estudaremos apenas um deles: o teste de Tukey-Kramer. Provavelmente, seja o teste de contrastes mais utilizado e relativamente fácil de entender.

No método de Tukey-Kramer, a diferença mínima significativa (DMS) é calculada para cada par de médias. Essa diferença depende do tamanho amostral de cada grupo, da variação média dentro dos grupos e do número total de grupos. Para um desenho balanceado, todas DMS serão as mesmas. Se a diferença observada entre um par de médias é maior do que a DMS, esse par de médias é significativamente diferente. No exemplo das abundâncias de aranhas nos tipos de florestas, os valores dos testes de Tukey-Kramer (abaixo da diagonal principal) e os valores de p (acima da diagonal) são apresentados na Tabela 14.

Tabela 14 – Valores do teste de Tukey-Kramer e valores de p

	Floresta nativa	Floresta de Araucária	Floresta de pinus	Floresta de eucaliptos
Floresta nativa (FN)	-	0,767	0,087	0,825
Floresta de Araucária (FA)	1,380	-	0,325	0,325
Floresta de pinus (FP)	3,976	2,596	-	0,026
Floresta de eucaliptos (FE)	1,216	2,596	5,192	-

Fonte: Baldissera *et al.* (2008a).

Como se pode verificar, somente as médias das abundâncias de aranhas na floresta de *Pinus* e na floresta de *Eucalyptus* foram diferentes entre si ($p = 0,026$). Também podemos olhar para os intervalos de confiança para verificar quais médias são diferentes. A Tabela 15 mostra o resultado do teste de contrastes com os intervalos de confiança para cada par de grupos. A coluna diferença mostra a diferença entre as duas médias, a segunda e terceira colunas mostram os limites e a quarta coluna os valores de p . Note que o único intervalo de confiança que não engloba o zero (igualdade de médias) é o do contraste entre Floresta de eucaliptos e Floresta de pinus.

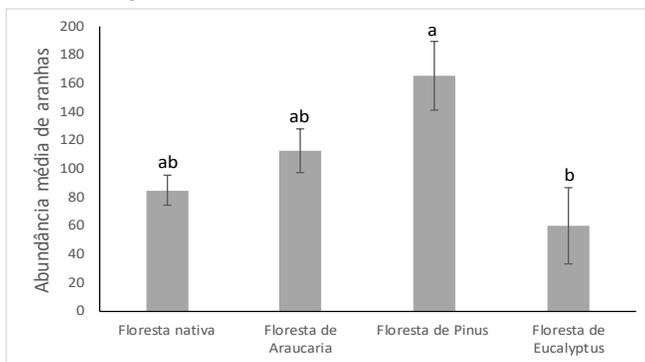
Tabela 15 – Resultado do teste de Tukey-Kramer com limites de 95% de confiança e valores de p

	Diferença	Limite inferior	Limite superior	Valor de p
FA × FN	28,00	-63,88	119,88	0,766
FE × FA	-24,67	-116,55	67,22	0,825
FP × FN	80,67	-11,22	172,55	0,086
FE × FA	-52,67	-144,55	39,22	0,325
FP × FA	52,67	-39,22	144,55	0,325
FP × FE	105,33	13,45	197,22	0,026

Fonte: Elaborada pelo autor (2021).

Para apresentar os resultados, pode ser uma boa opção fazer um gráfico de barras ou de caixas. O gráfico de barras é mais simples e mostra a média e o erro-padrão. Na Figura 17, as médias são indicadas pelas barras, os erros-padrões pelas linhas verticais e as letras indicam quais médias são diferentes entre si. Há outras formas de ilustrar os resultados de uma ANOVA. A escolha pode depender do tipo de gráfico que uma revista científica prefere, ou de como as pessoas na área de conhecimento do pesquisador normalmente apresentam esses dados.

Figura 17 – Médias \pm erros-padrões da abundância de aranhas arbustivas em 4 ambientes florestais



Fonte: Baldissera *et al.* (2008a).

6.3.4 ANOVA de duas vias

Também conhecida como ANOVA bifatorial, usa-se quando há uma variável quantitativa e duas variáveis nominais. Para cada observação, deve se ter um valor para cada uma das variáveis nominais. Para esse teste, temos três hipóteses nulas que são testadas: as médias agrupadas em um fator são iguais, as médias agrupadas no outro fator são iguais e não há interação entre os dois fatores. Quando os tamanhos amostrais em cada subgrupo são iguais (desenho balanceado), calcula-se o quadrado médio para cada um dos dois fatores (os efeitos principais), para a interação, e para a variação dentro de cada combinação de fatores. Então, calcula-se cada estatística F dividindo o quadrado médio pelo quadrado médio dentro de subgrupos. Se houver uma interação entre os fatores, a sugestão usual é que não se deve olhar os resultados isolados dos fatores. Por quê? Porque a variação em um fator é dependente da variação no outro fator. O que se pode fazer se a interação é significativa é olhar para cada fator separadamente, usando uma ANOVA de uma via.

Como todas as ANOVAs, a de duas vias assume que as observações são independentes, e que a variável aleatória quantitativa é normalmente distribuída e os grupos possuem mesmos desvios-padrões. Já abordei, na parte de ANOVA de uma via, todos os aspectos sobre esses pressupostos. Assim como naquela análise, esses pressupostos podem ser analisados graficamente ou com os testes de Shapiro-Wilk para normalidade e Bartlett para homocedasticidade.

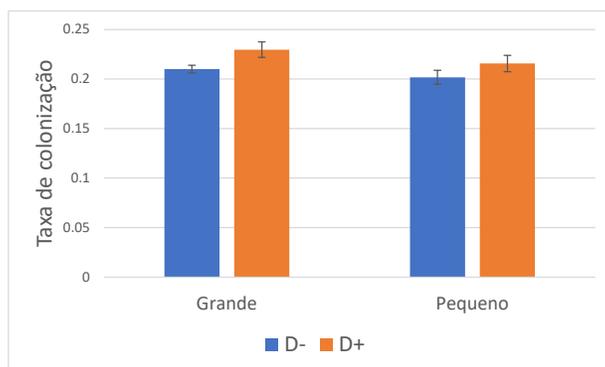
Assim como na ANOVA, podem ser utilizados limites de confiança para verificar as diferenças entre médias em testes de contraste caso haja efeito dos fatores e para saber em qual amplitude de valores podemos esperar uma média com 95% de confiança.

6.3.4.1 Exemplo de ANOVA de duas vias

Baldissera *et al.* (2013) realizaram um experimento de campo para testar se as taxas de colonização por aranhas em manchas de vegetação eram influenciadas pelo tamanho da mancha (grande/pequena) e pela diversidade vegetal (maior/menor). Então, temos dois fatores: tamanho da mancha e diversidade da mancha; e uma variável aleatória quantitativa taxa de colonização. Note que, agora, temos mais hipóteses que são testadas. No caso das taxas de colonização por aranhas, a hipótese de estudo era que manchas grandes e mais diversas teriam maior taxa de colonização. Portanto, os autores esperavam encontrar uma interação entre os dois fatores. Isso significa que a taxa de colonização aumentaria em manchas grandes e mais diversas ao mesmo tempo.

É importante notar que a inserção de mais uma variável nominal no teste pode tornar as coisas mais difíceis de discutir. No caso do exemplo acima, só há dois níveis de cada fator, então, na presença de interação os cruzamentos não são muito difíceis de visualizar. A Figura 18 mostra as médias e erros-padrões das taxas de colonização por aranhas para os dois fatores.

Figura 18 – Médias \pm desvios-padrões de taxas de colonização por aranhas em manchas grandes ou pequenas e mais ou menos diversas



Fonte: Baldissera *et al.* (2013).

Note que as taxas de colonização foram menores em manchas de vegetação pequenas (Pequeno) e menos diversas (D-), o que poderia corroborar a hipótese de estudo. A Tabela 16 mostra os resultados das análises.

Tabela 16 – Resultado da ANOVA de duas vias

Efeito	Soma de quadrados	Graus de liberdade	Quadrados médios	F	<i>p</i>
Diversidade	0,001	1	0,001	1,819	0,196
Tamanho	0,001	1	0,001	0,808	0,382
Diversidade \times Tamanho	0,000	1	0,000	0,052	0,822
Total	0,012	16	0,001		

Fonte: Baldissera *et al.* (2013).

Como se pode ver, não houve nenhum efeito significativo (todos valores de $p > 0,05$). Se houvesse interação entre os fatores, poderia se testar a diferença na média de colonização em relação ao tamanho das manchas e, depois, em relação à diversidade.

6.3.5 ANOVA aninhada

Utiliza-se a ANOVA aninhada (também conhecida como anova hierárquica ou multinível) quando se tem uma variável quantitativa e duas ou mais variáveis nominais. A diferença em relação à ANOVA de duas vias é que as variáveis nominais estão aninhadas. Isso significa que cada valor de uma variável

nominal em um nível inferior (os subgrupos) é encontrado em combinação com apenas um valor de uma variável nominal de um nível superior (os grupos).

Então, essa análise é uma extensão da ANOVA de uma via, na qual cada grupo é dividido em subgrupos. Em teoria, se escolhem esses subgrupos aleatoriamente de um conjunto maior. A lógica dos subgrupos é que eles estão inseridos em um grupo e, portanto, teoricamente, não há independência nas observações. Um exemplo das ciências sociais pode esclarecer isso de forma bem simples. Imagine que um pesquisador esteja interessado em avaliar a taxa de aprendizagem da disciplina de biologia de estudantes de turmas do Ensino Médio de duas escolas. Suponhamos que haja 3 turmas com professores diferentes em cada escola. Se o pesquisador avaliar a aprendizagem em um estudante de cada professor, poderia comparar as diferenças entre as médias das duas escolas com uma ANOVA de uma via.

Entretanto, claramente, esse desenho experimental estaria abdicando de uma fonte importante de variação, o professor. Nesse caso, para um estudo mais completo da variação na aprendizagem, pode-se selecionar, aleatoriamente, 10 estudantes por professor. Nesse caso, cada estudante estaria associado a apenas um professor. Dessa forma, o desenho amostral tem 2 níveis (variáveis nominais): professores e escolas. Nitidamente, há uma dependência na aprendizagem entre os estudantes do mesmo professor, que, por sua vez, tem dependência com a escola. Um exemplo biológico desse tipo de desenho foi dado anteriormente. É o das 3 galhas em cada galho de 20 arbustos em 4 sítios. Nesse caso, temos 3 variáveis nominais aninhadas (sítios/arbustos/galhos).

Quando se faz uma ANOVA aninhada, em geral estamos interessados em testar uma hipótese nula sobre as médias dos grupos, não importando se os subgrupos são significativamente diferentes. Por essa razão, pode-se ficar tentado a ignorar os subgrupos e usar todas as observações em uma ANOVA de uma via. Isso seria um erro. A independência das observações é um dos pressupostos de uma ANOVA. No caso dos estudantes, estaríamos tratando todos eles como observações independentes dentro da mesma escola. Apesar de padronizações nos processos de ensino-aprendizagem, cada professor tem sua própria dinâmica, a qual pode influenciar na aprendizagem. Fazendo isso em uma ANOVA de uma via, estaríamos comparando a diferença nas médias de grupos em relação à variação dentro de cada grupo, fingindo que há muitas medidas independentes de aprendizagem (10 estudantes \times 3 professores = 30 observações). Dessa forma, teríamos violado a pressuposição de independência da ANOVA de uma via e estaríamos usando de pseudorrepetição (HURLBERT, 1984).

Uma ANOVA aninhada possui uma hipótese nula para cada nível. Em uma ANOVA aninhada de dois níveis, uma hipótese nula é que os grupos possuem a mesma média. Para nossas escolas, essa hipótese seria que os estudantes da escola 1 possuem a mesma média de aprendizagem do que a escola 2. A segunda hipótese nula é que os subgrupos dentro de cada grupo possuem as mesmas médias. Para o exemplo, essa hipótese seria que todos os estudantes dos 3 professores da escola 1 possuem a mesma média e que todos os estudantes dos 3 professores da escola 2 também possuem a mesma média (que poderia ser diferente da média dos estudantes da escola 1).

6.3.5.1 Exemplo de ANOVA aninhada

Uma anova aninhada foi usada para avaliar a resposta da diversidade de besouros de solo em zonas ripárias em uma unidade de conservação no estado do Paraná (REBESCHINI, 2021). Três zonas ripárias foram selecionadas em riachos com matas ciliares e três zonas ripárias em riachos sem matas ciliares. Dentro de cada zona ripária, foram selecionadas três unidades amostrais em três distâncias do riacho: 5 m, 25 m e 45 m. A lógica é que há um gradiente de alteração ambiental conforme se distancia do corpo d'água que pode influenciar na resposta dos besouros. Nesse desenho amostral, as distâncias estão aninhadas dentro dos tipos de riachos. A técnica de coleta utilizada foi a armadilha de queda, que consiste em um copo enterrado no solo com uma mistura de álcool e água, que permanece aberto durante uma semana, durante a qual os bichos caem nele. Cada copo é, portanto, uma unidade amostral. A variável resposta foi o número de espécies de besouros encontrado em cada ponto (copo) e as variáveis nominais são tipo de zona ripária (2 níveis) e distância ao riacho (3 níveis).

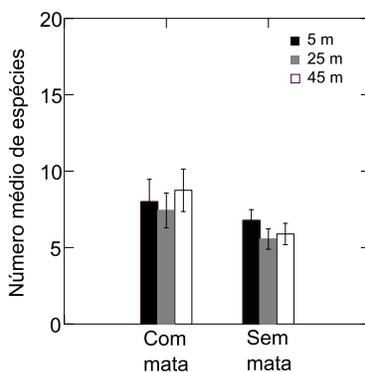
Tabela 17 – Resultado da ANOVA aninhada

Efeito	Soma de quadrados	Graus de liberdade	Quadrados médios	F	p
Zona ripária	66,7	1	66,67	7,978	0,007
Distância (Zona ripária)	26,4	4	6,59	0,789	0,538
Total	401,1	48	8,36		

Fonte: Rebeschini *et al.* (2021).

A Tabela 17 mostra que houve uma diferença significativa nas médias dos números de espécies de besouros encontrados nos dois tipos de zonas ripárias ($p = 0,017$), mas não houve qualquer efeito da distância ao riacho ($p = 0,790$). A zona ripária com mata ciliar teve um número médio de espécies de besouros de 8,08 ($\pm 0,72$ erro-padrão), enquanto a zona ripária sem mata ciliar teve $6,07 \pm 0,37$. A diferença entre as médias foi -2,22. O intervalo de confiança foi -3,87 a -0,65.

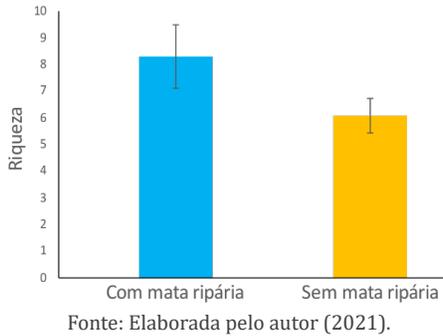
Figura 19 – Médias \pm erros-padrões do número de espécies de besouros em três distâncias de riachos



Fonte: Rebeschini (2021).

A Figura 19 mostra as médias dentro de cada subgrupo e, também, ilustra a diferença nas médias entre os grupos, apesar de não mostrar as médias dos grupos explicitamente. A maneira de fazer o gráfico dos resultados de uma ANOVA aninhada depende do resultado e da questão de estudo. Se a variação entre os subgrupos é interessante, pode-se mostrar as médias de cada subgrupo com diferentes padrões de cores, indicando os diferentes grupos, como mostrado na Figura 19. Se a variação entre subgrupos não é significativa e a variação entre grupos é significativa – o interesse é somente nos grupos e uma ANOVA aninhada foi usada para ver se era certo combinar os subgrupos –, deve-se apenas mostrar, como apresenta a Figura 20, as médias dos grupos em um gráfico de barras.

Figura 20 – Médias \pm erros-padrões do número de espécies de besouros em dois tipos de matas ripárias

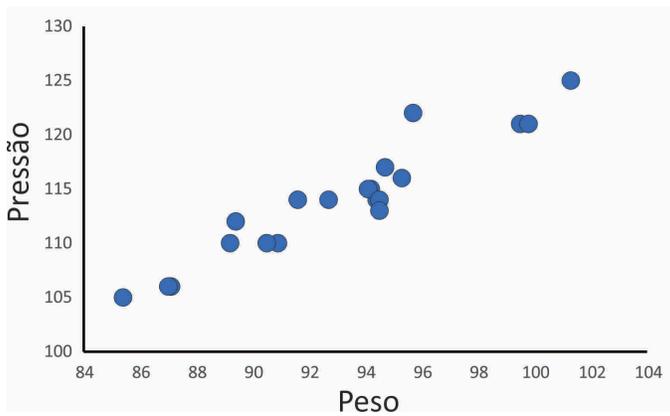


Quando os tamanhos amostrais em uma ANOVA aninhada são diferentes, os valores de p dos testes F podem não ser uma boa estimativa da probabilidade real. Por isso, deve-se procurar sempre realizar desenhos balanceados, seja em que tipo de ANOVA se estiver interessado. A ANOVA aninhada, como todas as ANOVAs, assume que as observações dentro de cada subgrupo são normalmente distribuídas e possuem desvios-padrões iguais.

6.3.6 Correlação e regressão linear simples

Um dos gráficos mais comuns na ciência é o que mostra uma variável quantitativa no eixo X (horizontal) e outra variável quantitativa no eixo Y (vertical). Há basicamente duas análises que são realizadas para associar duas variáveis quantitativas: correlação e regressão. Essas duas análises têm o mesmo objetivo: testar a associação entre duas variáveis quantitativas. Há algumas diferenças entre elas que são abordadas de formas diferentes em livros de estatística. A Figura 21 mostra um exemplo de gráfico de associação com dados retirados da Tabela 18, que mostra os pesos e as pressões sanguíneas de 20 indivíduos.

Figura 21 – Relação entre peso e pressão sanguínea



Fonte: Modificado de Penn State University (2018).

Tabela 18 – Valores de pressão sanguínea e peso

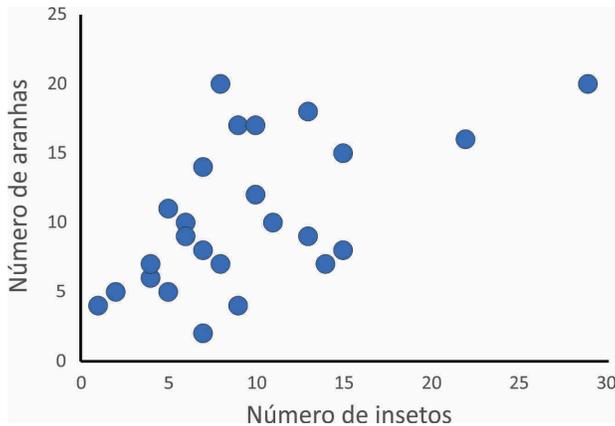
Indivíduo	Pressão	Peso
1	105	85,4
2	115	94,2
3	116	95,3
4	117	94,7
5	112	89,4
6	121	99,5
7	121	99,8
8	110	90,9
9	110	89,2
10	114	92,7
11	114	94,4
12	115	94,1
13	114	91,6
14	106	87,1
15	125	101,3
16	114	94,5
17	106	87
18	113	94,5
19	110	90,5
20	122	95,7

Fonte: Modificado de Penn State University (2018).

Há basicamente três objetivos para as análises com esses dados. O primeiro é realizar um teste de hipótese para saber se há uma associação estatística significativa entre peso corporal e pressão sanguínea. Em outras palavras, conforme aumenta a variável X , há uma evidência significativa de mudança na variável Y (para mais ou para menos)? No caso citado, pode-se querer saber se a pressão sanguínea foi significativamente maior em indivíduos mais pesados. Mas essa relação está tão óbvia no gráfico que um teste desses não é interessante. Note que a distribuição dos objetos no gráfico forma um padrão muito parecido com uma linha. O valor de p para a relação acima é menor do que 0,0001, ou seja, há uma forte evidência de que há uma associação significativa entre as variáveis. A probabilidade de esta associação ser devida simplesmente ao acaso é menor do que 0,01%.

Agora, considere a Figura 22, que mostra a relação entre o número de insetos e o número de aranhas em uma floresta. A relação não é tão óbvia quanto a do gráfico anterior. Com algum esforço podemos perceber que há mais aranhas onde há mais insetos. Porém, a nuvem de pontos está mais espalhada comparada com a do gráfico anterior. Apesar disso, há uma relação significativa entre as duas variáveis ($p = 0,003$).

Figura 22 – Relação entre número de insetos e número de aranhas



Fonte: Baldissera *et al.* (2008b).

O segundo objetivo das análises é descrever o quão forte é a associação entre as variáveis. Note que o valor de p mostra se há ou não uma associação significativa, mas não diz nada sobre a força de associação. Isso é geralmente expresso por duas estatísticas: o coeficiente de correlação de Pearson (r) e o coeficiente de determinação (r^2). O primeiro é utilizado na análise de correlação e varia de -1 até +1. O segundo é usado na regressão e varia entre 0 e 1. Para os dados de pressão sanguínea, há uma relação muito forte, como mostrada pelo valor de $r^2 = 0,903$. Isso significa que em torno de 90% da variação na pressão sanguínea pode ser explicada pela variação no peso. O valor de r^2 para os dados das aranhas de teia é 0,329. Ou seja, somente em torno de 33% da variação é explicada. Isso quer dizer que há 67% de variação no número de aranhas na floresta, a qual está associada com outros efeitos.

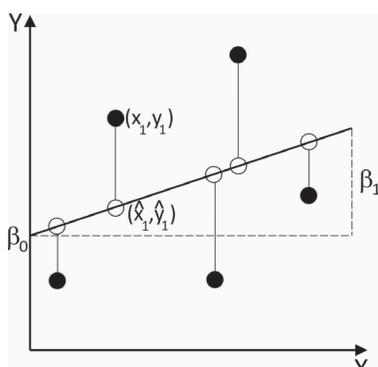
O último objetivo das análises é descrever a relação entre a variável resposta e a variável explicativa para prever ou predizer os valores de y para um dado valor de x . Para isso, se utiliza um modelo que é uma descrição matemática

simplificada de um sistema natural real. Esse modelo é calculado em uma análise de regressão linear simples. Essa análise é usada para calcular parâmetros de uma equação de primeiro grau relacionando as variáveis Y e X . A regressão linear é chamada simples quando se aplica a casos em que há apenas uma variável explicativa. Neste caso, podemos pensar que Y muda em função de X , ou seja, $Y = f(x)$. A equação da reta é dada pela fórmula

$$Y = \beta_0 + \beta_1 X$$

que corresponde à equação de uma linha reta (daí o linear) que cruza a nuvem de pontos de uma forma ótima e permite o cálculo de um valor estimado de y para cada valor de x (LEGENDRE; LEGENDRE; 2012). β_0 é um coeficiente que estima o intercepto da reta de regressão com o eixo Y . β_1 é a inclinação da reta de regressão, também chamado coeficiente da regressão. A Figura 23 mostra cinco coordenadas x, y correspondendo a 10 medidas de duas variáveis (círculos cheios) e os valores estimados para a reta ajustada (círculos vazados) e os valores dos coeficientes. A distância entre cada observação e a reta ajustada ao modelo é chamada de resíduo e um dos métodos utilizado para calcular a equação da reta é o que minimiza a soma de quadrados destes resíduos, oferecendo o melhor ajuste. Esse método é chamado de quadrados mínimos.

Figura 23 – Ajuste de uma reta de regressão



Fonte: Elaborada pelo autor (2021).

Para os dados de pressão sanguínea, a equação é $Y = 7,407 + 0,752X$. Isso quer dizer que a pressão sanguínea seria 7,407 (intercepto) quando o peso fosse zero, o que não é real, obviamente. No gráfico, esse valor está oculto, pois o eixo X iniciou no valor de 85, que é o menor peso de um indivíduo da amostra. Alternativamente, para que o intercepto possa ser melhor interpretado, pode-se centralizar a variável X na média amostral e recalculer o coeficiente. Isso é particularmente interessante quando o intercepto calculado sem a centralização carece de um significado biológico. No caso do exemplo, o novo intercepto, após a centralização, foi de 2,205. Por outro lado, a pressão sanguínea aumenta 0,752 para cada 1 kg de aumento de peso. Essa é provavelmente a parte mais útil da análise para esses dados, pois pode-se prever, com certa precisão, a pressão de um indivíduo a partir do seu peso. Para os dados das aranhas de teia a equação é $Y = 5,719 + 0,492X$. Nesse caso, qualquer previsão é pouco válida e apenas saber que há significativamente mais aranhas quanto mais insetos houver (o teste de hipótese) é mais interessante do que saber a equação da linha.

Há uma variável nominal que mantém as duas variáveis quantitativas em pares. No caso da pressão sanguínea é o número de cada indivíduo. Isso é importante, pois as medidas das duas variáveis quantitativas devem ser realizadas na mesma unidade amostral, seja ela um indivíduo, uma área, um volume, uma cidade, uma região, etc. Essa coluna não entra na análise e sua função é identificar a lista de objetos.

Em alguns casos, testar uma associação entre duas variáveis quantitativas se direciona a uma questão biológica sobre relação de causa e efeito entre as variáveis X e Y . Um exemplo é ministrar às pessoas quantidades diferentes de uma droga e medir sua pressão sanguínea. A hipótese nula é que não há relação entre a quantidade da droga e a pressão sanguínea. Se você rejeitar a hipótese nula, concluirá que a quantidade de droga causa mudanças na pressão sanguínea. Esse tipo de experimento, em geral, é realizado quando se pode ter controle sobre os objetos de estudo ou define-se, antes de realizar o experimento, que determinada variável de interesse é causa de outra. No exemplo da pressão sanguínea, o peso é a variável que causa a mudança na pressão. Outro exemplo, aranhas tecedoras de teias ocupam arbustos de florestas e se alimentam majoritariamente de insetos. Quanto maior o número de insetos para as aranhas, maior o número de espécies de aranhas que a floresta poderá conter.

Em outros casos, se quer saber se duas variáveis estão associadas sem necessariamente inferir uma relação de causa e efeito. Nesse caso, não se classifica antes qualquer variável. As duas variam naturalmente. Se houver uma as-

sociação entre elas, infere-se que a variação em X pode causar a variação em Y , ou a variação em Y pode causar a variação em X , ou uma variação em algum outro fator pode afetar tanto Y quanto X . Um exemplo é medir a quantidade de uma proteína na superfície de algumas células e o pH do citoplasma dessas células. Se a quantidade de proteína e o pH forem correlacionados, pode ser que a quantidade de proteína afeta o pH interno, ou o pH interno afeta a quantidade de proteína. Ou, ainda, há algum fator, tal como a concentração de oxigênio (não medida), que afeta tanto a concentração da proteína e o pH. A associação de aranhas de teia com os insetos é um exemplo disso. O número de insetos influencia até certo ponto o número de aranhas (33% para ser específico). Então, há uma variação no número de aranhas que não é associada com o número de insetos. No caso da pressão sanguínea, a força de associação é muito forte, mas, no caso das aranhas e dos insetos, a força é bem menor. Com a equação de uma linha que se ajusta à nuvem de pontos pode-se, então, fazer previsões. Nesse tipo de análise, o pesquisador começa a se aprofundar no universo da modelagem de dados. Um modelo é uma representação de como um ou mais processos influenciam no funcionamento de um sistema.

As ferramentas estatísticas usadas para o teste de hipótese, descrevendo a intimidade da associação, e o desenho de uma linha por meio dos pontos, são a correlação e a regressão. Correlação e regressão linear fornecem o mesmo valor de p para o teste de hipótese e, para muitos experimentos, isso pode ser o que realmente importa. Então, se houver interesse principalmente no valor de p , a diferença entre correlação e regressão não é importante.

A principal diferença entre a correlação e a regressão é que na correlação se amostram as duas variáveis quantitativas aleatoriamente de uma população sem preocupação em uma possível relação de causa-efeito. Na regressão, há uma ideia clara de qual variável é explicativa (X). Entretanto, tentar classificar cada experimento como regressão ou correlação pode ser confuso. Dessa forma, é importante prestar atenção ao objetivo do experimento. É sempre bom ficar ciente que uma análise de regressão com dados coletados em experimentos naturais, em geral, não se prestará para realizar previsões, pois haverá muitas variáveis que não serão controladas (variáveis de confusão). Em geral, o r^2 ou o r será pequeno e os pontos no gráfico estarão em uma nuvem dispersa como o do exemplo das aranhas de teia e os insetos. Nesse experimento, especificamente, bastaria saber que o número de aranhas está associado ao número de insetos (o teste de hipótese). Determinar uma equação para a reta de regressão não seria adequado, uma vez que a força da associação foi baixa. Então, se houver

interesse principalmente em utilizar o valor de p para testar uma hipótese, para ver se há uma associação significativa entre duas variáveis, não importa muito se você usar o teste de regressão ou correlação.

Uma coisa que se ouve bastante entre pesquisadores é que “correlação não implica causa.” Isso é um lembrete de que, quando se está amostrando a variação natural de duas variáveis A e B, há também variação natural em várias outras variáveis de confusão que podem causar a associação entre A e B. Já abordamos esse contratempo quando estudamos sobre as variáveis de confusão. Então, se há uma associação significativa entre A e B, isso não significa necessariamente que a variação em A é a mais importante para causar a variação em B. Pode haver alguma outra variável, C, que afeta as outras duas. Isso é especialmente verdade em experimentos que não controlam variáveis de confusão, como experimentos de campo ou observacionais. Desde que é possível pensar em múltiplas explicações para a associação entre duas variáveis, isso significa que se deve sorrir cinicamente, dizer que “correlação não implica em causa” e descartar qualquer estudo de correlação da variação natural que ocorre nos sistemas? Não. A observação da correlação entre duas variáveis pode sugerir que há algo interessante ocorrendo, algo que se pode investigar mais a fundo.

A hipótese nula da correlação/regressão linear é que a associação é igual a zero. Em outras palavras, conforme a variável X fica maior, a variável Y não fica nem maior, nem menor. Um intervalo de confiança para a correlação pode ser usado para decidir aceitar ou rejeitar a hipótese nula. Se o zero (sem correlação) está contido no intervalo, a hipótese nula é aceita. Para uma análise de regressão linear simples, pode-se calcular três intervalos de confiança. O primeiro é o do coeficiente β_0 . O segundo é o do coeficiente β_1 . O último é o de cada valor do ajuste da reta de regressão. Esse intervalo de confiança dá uma ideia da amplitude de variação em que se espera encontrar a medida “real” do valor ajustado com 95% de confiança. Com os intervalos de confiança de todos os valores ajustados pode-se traçar linhas no gráfico de regressão para visualizar quais valores observados estão dentro ou fora dos limites de confiança. Para a análise da associação entre peso e pressão arterial, o intervalo de 95% de confiança para β_0 foi -15,99 a 20,4 e para β_1 foi 1,01 a 1,39. Isso indica que, se forem realizados 100 experimentos a partir da reamostragem desse universo amostral, podemos esperar que, em 95% deles, os coeficientes “reais” estejam dentro dos limites de confiança.

Os pressupostos mais importantes para as análises são, como de costume, a independência entre as observações, a normalidade e a homocedasticidade,

além da linearidade. A linearidade diz respeito ao padrão da relação entre as duas variáveis, que deve ser linear. O padrão da relação pode ser positivo ou negativo, mas o importante é que a nuvem de pontos formada no gráfico de relação entre as duas variáveis mostre que se pode desenhar uma linha reta por meio dela e não uma linha curva. Por exemplo, os gráficos das figuras 22 e 23 mostram padrões de linearidade. A adequação dos dados aos pressupostos pode ser analisada de forma gráfica ou estatística. A normalidade é testada nos valores dos resíduos, que são os desvios dos valores observados em relação aos valores estimados. Simulações mostraram que efeitos de desvios moderados da normalidade não influenciam no poder dos testes de correlação e regressão (EDGELL; NOON, 1984).

Em geral, quando há homocedasticidade, há uniformidade no espalhamento dos pontos em torno da reta de regressão. Se houver um ou mais pontos no gráfico XY que estão afastados da reta de regressão ou do conjunto maior de pontos, teremos o que se chama de *outliers*, ou pontos fora da curva. Uma forma gráfica de analisar pontos fora da curva é montando um gráfico dos resíduos da regressão no eixo Y com os valores ajustados no eixo X . Dados homocedásticos e sem pontos fora da curva mostrarão uma dispersão equitativa de valores dos resíduos acima e abaixo do valor zero. Qualquer ponto extremo deve ser analisado. Dados fora da curva podem ser resultados de erros de medida, de erros na hora da transcrição dos valores para uma planilha de dados, ou são próprios do sistema em estudo. É importante ressaltar que, nem sempre, valores fora da curva são inúteis. Em geral, estudantes costumam ficar desanimados quando surge um ou mais pontos fora da curva. Às vezes, esses valores, se não forem resultado de erros de medida ou registro, podem informar que há algo interessante acontecendo. Por exemplo, suponha que sejam realizadas medidas de um indicador de poluição em 20 riachos. Se todas as medidas foram cuidadosamente realizadas e transcritas e aparecer um ponto fora da curva com altos valores de poluição, essa informação será extremamente relevante.

Quando há pontos fora da curva, uma alternativa é utilizar a correlação de Spearman, que reduz a influência de valores extremos, ou achar uma transformação de dados que torne os dados mais homogêneos. Outra abordagem é analisar os dados sem os valores extremos e reportar os resultados com e sem esses valores extremos.

A forma tradicional de apresentar os resultados de uma regressão simples é com um gráfico de dispersão com a variável independente no eixo X e a variável dependente no eixo Y . No caso de uma correlação que não assume causa-efeito entre as variáveis, tanto faz apresentar uma ou outra variável em um ou outro eixo.

6.3.7 Regressão linear múltipla

A regressão linear múltipla é uma extensão lógica da regressão simples de uma para muitas variáveis explicativas quantitativas. Então, quando há uma variável resposta e duas ou mais variáveis explicativas, utiliza-se a regressão múltipla. O modelo básico da regressão linear múltipla é o seguinte:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

Em comparação com o modelo de regressão simples, há a adição de novas variáveis independentes X_n e seus coeficientes β_n . Os propósitos da regressão múltipla são basicamente dois: (1) prever uma resposta desconhecida da variável Y baseada nas variáveis X_n e (2) tentar entender as relações funcionais entre a variável resposta e as variáveis explicativas. Entretanto, deve-se ter muito cuidado ao realizar uma análise de regressão linear múltipla. O simples fato de adicionar mais variáveis explicativas a um modelo de resposta da variável Y já aumenta a probabilidade de encontrar algum tipo de influência significativa, pois é óbvio que a variação em Y é causada por alguma(s) variável(is)! Dessa forma, pode ser muito tentador simplesmente sair por aí coletando um monte de variáveis que julgamos importantes para explicar a variação em Y . Não é aconselhável fazer isso. Se houver um objetivo e hipóteses claras para o estudo, essa armadilha será evitada.

A ideia básica dessa análise, então, é encontrar uma equação matemática que forneça uma relação linear entre as variáveis X e a variável Y . O coeficiente β_1 é a inclinação da curva estimada da regressão de Y em X_1 , se todas as outras variáveis permanecerem constantes e, assim por diante, para as outras variáveis; β_0 é o intercepto. Resumidamente, a análise encontra os valores de $\beta_1, \beta_2,$ etc. – chamados de coeficientes parciais de regressão – e de β_0 que minimizam os desvios ao quadrado entre os valores esperados e observados (que definem a linha de regressão) de Y . Os desvios são os resíduos como vimos anteriormente na regressão linear simples.

A estatística que mostra o quanto a equação se ajusta aos dados é o r^2 (coeficiente de determinação múltipla), que varia de 0 (sem associação entre Y e X_n) a 1 (associação perfeita). O valor de p é uma função de r^2 , do número de observações (objetos) e do número de variáveis X . Como ressaltado antes, o aumento do número de variáveis pode influenciar na probabilidade de se encontrar um resultado significativo. Uma forma de diminuir esse viés é utilizar a estatística r^2

ajustado (BLANCHET *et al.*, 2008), que corrige o valor de r^2 baseado no número de variáveis X . De qualquer forma, é importante que se saiba qual estatística se utiliza na área do conhecimento do pesquisador, pois apresentar uma outra estatística, mesmo que possa ser considerada mais acurada, pode ser estranho para outros pesquisadores e será necessário justificar muito bem a escolha.

A lógica de ajustar o r^2 é que quanto mais variáveis inúteis se acrescentam para o modelo de regressão, menor o valor do r^2 ajustado; e quanto mais variáveis úteis se acrescentam, maior esse valor. A principal diferença é que o r^2 sem ajuste assume que cada variável X explica uma parte da variação em Y , enquanto o r^2 ajustado mostra a porcentagem da variação explicada pelas variáveis X que realmente afetam a variável Y .

Há dois tipos de coeficientes resultantes da regressão múltipla: coeficiente parcial e coeficiente parcial padronizado. O primeiro é utilizado quando se quer prever a resposta da variável Y com valores estimados das variáveis X . Esse coeficiente não dá informação sobre a importância relativa de cada variável. O segundo coeficiente é utilizado quando se quer estudar as relações funcionais. A magnitude do coeficiente parcial de regressão padronizado indica a importância relativa de cada variável, então aquelas variáveis com maiores valores do coeficiente padronizado possuem uma relação mais forte com a variável Y . Assim como para a regressão linear simples, para regressão linear múltipla é possível calcular os intervalos de confiança para o coeficiente β_0 e todos os coeficientes β_n das variáveis explicativas.

Assim como a regressão linear simples, a regressão linear múltipla assume que as variáveis são normalmente distribuídas, que há homocedasticidade e que cada variável independente é função linear da variável dependente. A última pressuposição é especialmente difícil de testar. É importante olhar as correlações entre a variável Y e todas as outras variáveis, mas aparentar uma relação linear nem sempre será o suficiente para assumir esse tipo de relação. Um pressuposto novo e importante da regressão múltipla é que não há multicolinearidade, ou seja, que as variáveis independentes não são altamente correlacionadas entre si. Altamente correlacionada é meio abstrato. Será que um r de 0,5 é suficiente para duas variáveis serem “altamente” correlacionadas? Em geral, assume-se que uma correlação entre $\pm 0,39$ a $\pm 0,69$ é forte, enquanto maiores do que $\pm 0,7$ são muito fortes. Mas, por que isso é importante? Como a regressão múltipla calcula coeficientes parciais para cada variável explicativa, caso duas variáveis sejam correlacionadas, há uma mistura de efeitos e o coeficiente parcial de uma variável tem uma variação intrínseca relacionada com a outra variável da qual

ela é correlacionada. Em outras palavras, se uma variável explicativa tiver um efeito significativo sobre a variável resposta, pode ser que esse efeito seja devido a uma variação na outra variável explicativa. Então, antes de realizar uma regressão múltipla, é uma boa prática montar uma matriz de correlação entre todas as variáveis explicativas. Se houver correlações fortes ou muito fortes e significativas, a sugestão é escolher uma das variáveis para inserir no modelo. A chave para interpretar os resultados é que, se essa variável tiver um efeito significativo sobre Y , então já se sabe que as variáveis correlacionadas com ela que foram deixadas de fora do modelo também são importantes para explicar esse efeito.

6.3.7.1 Seleção de variáveis

Independente do objetivo da regressão linear múltipla, se há muitas variáveis X , algumas delas terão mais força de relação com a variável Y e outras terão menos força. Dessa forma, pode-se querer saber quais variáveis são mais ou menos importantes. Isso determinará qual o modelo reduzido final a ser utilizado. O princípio da parcimônia, formulado por Guilherme de Ockham, no século 13, declara que a “multiplicidade não deve ser inserida sem necessidade”. Em outras palavras, variáveis desnecessárias devem ser evitadas. Tendo isso em mente, pode-se realizar uma análise de seleção de variáveis para diminuir a complexidade do modelo completo inicial e descartar variáveis desnecessárias. Um método para a regressão linear múltipla é o de escolha gradativa (*stepwise selection*), no qual há duas opções: análise para frente ou análise para trás. Nos dois casos, será preciso decidir quando parar a análise. A orientação é sempre utilizar o valor de r^2 ajustado e um valor de p estipulado (que pode ou não ser o tradicional 0,05). Se a adição ou subtração de uma variável no modelo não aumenta significativamente o valor de r^2 ajustado, sua inclusão ou subtração não faz diferença na resposta de Y . Na análise para frente, inicia-se com a variável X com maior valor de r^2 ajustado e vai se adicionando as outras variáveis explicativas uma a uma. Se há um aumento significativo em r^2 ajustado, a nova variável é inserida no modelo até que mais nenhuma influencie significativamente. Um método para frente com uso do r^2 ajustado foi descrito por Blanchet *et al.* (2008) e chamado de seleção de variáveis para frente ajustado, que será utilizado no exemplo do teste. Na análise para trás, inicia-se com o modelo total e vai se retirando as variáveis conforme elas vão apresentando menor contribuição para o valor de r^2 ajustado e, conseqüentemente, o valor de p é menor do que o estipulado.

Os passos citados podem ser realizados manualmente pelo pesquisador. Porém, o mais correto é realizar todas as possíveis combinações de variáveis explicativas em cada passo dos dois processos. Isso é particularmente tedioso quando se tem 3 ou mais variáveis. Os programas estatísticos geralmente realizam o que se chama regressão múltipla gradativa, na qual, cada vez que se adiciona ou subtrai uma variável do modelo, se testam os efeitos da remoção ou adição de todas as outras variáveis. Esse processo é realizado até que nenhuma adição ou subtração de uma variável X aumente significativamente o valor de r^2 ajustado.

Além do método para frente ajustado, outro método frequentemente utilizado é o de Akaike (AKAIKE, 1973), que gera um valor, chamado de critério de informação de Akaike, em inglês *AIC*, que pode ser usado para determinar, entre vários modelos, qual o mais provável de ser o melhor para um conjunto de dados. De qualquer forma, os dois métodos de seleção de variáveis apresentados levam em consideração a complexidade do modelo de regressão, ou seja, nas suas fórmulas é considerada a informação da quantidade de variáveis explicativas. Quanto mais variáveis explicativas, maior a probabilidade de encontrar um efeito. Porém, o modelo se torna mais complexo, e muitas variáveis podem ter uma pequena influência no efeito final, enquanto uma ou poucas variáveis podem ter uma influência decisiva. Não nos aprofundaremos nas lógicas analíticas desses dois métodos. A seleção de variáveis em uma regressão linear múltipla é um assunto delicado e controverso. Então, é fundamental que se pense muito bem no desenho amostral para reduzir ao máximo as variáveis explicativas dependendo do tamanho amostral. Também é importante que se observe como os pesquisadores na área de conhecimento costumam usar a regressão linear múltipla.

6.3.7.2 Exemplo de regressão linear múltipla

A Tabela 19 mostra o resultado de uma pesquisa de comunidades de peixes em 12 riachos no sul do Brasil. Os peixes foram coletados em unidades experimentais de 40 m de comprimento amostradas por 50 minutos em cada riacho. Para cada comunidade biológica, foi calculado um índice de diversidade funcional, baseado em características morfológicas de cada peixe ligadas à alimentação. Quanto maior o valor do índice, maior a diversidade de funções ligadas à alimentação que os peixes executam na comunidade (riacho). Em cada unidade amostral também foram coletadas seis informações físico-químicas da água: turbidez, alcalinidade, oxigênio dissolvido, condutividade elétrica, pH e temperatura.

Tabela 19 – Diversidade funcional de peixes e composição físico-química dos riachos

Riachos	DF	TURB	ALCA	OD	COND	TEM	PH
01	0,03	4,11	0,43	8,15	13,8	22	6,95
02	0,30	7,92	0,56	9,02	27,8	15	8,8
03	0,41	2,06	0,6	7,5	39,9	19	8
04	0,24	5,90	0,46	7,5	43,8	19	6,7
05	0,13	7,04	0,33	6,95	23,1	24	7,06
06	0,22	8,80	0,43	6,75	29,7	22	7,17
07	0,46	2,15	0,63	7	34,9	20	6,5
08	0,074	6,71	0,33	8,14	10,8	22	6,89
09	0,28	29,55	0,46	5,3	70,7	20	7,1
10	0,35	6,46	0,73	6,6	432,43	18	7,36
11	0,26	8,09	0,66	6	46	24	6,8
12	0,34	6,78	0,66	6,97	57,8	22	7,27

Fonte: Elaborada pelo autor (2021).

Sete riachos estavam em áreas preservadas, três em áreas impactadas e dois em áreas alteradas. Então, um objetivo é prever a diversidade funcional de peixes a partir das variáveis ambientais; enquanto o outro é gerar hipóteses sobre as causas da variação na diversidade funcional de peixes. Antes de realizar a análise de regressão múltipla, é importante analisar as correlações entre as variáveis explicativas para verificar a existência de multicolinearidade. Isso pode ser feito com uma matriz de correlação. A matriz dos índices de correlação de Pearson com os valores de p corrigidos é demonstrada na Tabela 20.

Tabela 20 – Matriz dos índices de correlação de Pearson entre seis variáveis explicativas

	TURB	ALCA	OD	COND	TEM	PH
TURB		1,000	0,660	1,000	1,000	1,000
ALCA	-0,195		1,000	0,990	1,000	1,000
OD	-0,589	-0,223		1,000	1,000	1,000
COND	0,024	0,546	-0,277		1,000	1,000
TEM	-0,014	-0,343	-0,403	-0,323		
PH	-0,016	0,199	0,468	0,067	-0,664	0,279

Fonte: Elaborada pelo autor (2021).

A Tabela 20 apresenta, na parte inferior, os valores dos índices de correlação de Pearson (r) das variáveis par a par; e, na parte superior, os valores de p com a correção de Bonferroni. Como não há nenhuma correlação significativa, podemos utilizar todas as variáveis para a análise de regressão linear múltipla. Uma regressão linear múltipla com todas as variáveis resulta em um r^2 de 0,91, um r^2 ajustado de 0,80, um valor de $p = 0,018$ e a equação da reta é:

$$\text{FD} = 7,406 - 0,012(\text{TURB}) + 0,307(\text{ALCA}) - 0,148(\text{OD}) - 0,0005(\text{COND}) - 0,045(\text{TEM}) + 0,026(\text{PH})$$

O valor 7,406 é o intercepto β_0 e os outros valores são os coeficientes para cada variável. Em resumo, o valor da diversidade funcional de peixes nos riachos varia negativamente em função das variáveis TURB, OD, COND e TEM; e, positivamente, em função das variáveis ALCA e PH. Com essa equação, podemos prever a variação no valor da diversidade funcional associada com alterações nos valores das variáveis. Por exemplo, o pesquisador vai até um outro riacho, mede as variáveis turbidez, alcalinidade, oxigênio dissolvido, condutividade, temperatura e pH e substitui esses valores na equação. O valor de diversidade funcional resultante é a previsão da diversidade funcional de peixes para aquele riacho.

Na análise anterior, não estávamos preocupados com as relações funcionais entre as variáveis explicativas e a variável resposta. Simplesmente, geramos uma equação que prevê o comportamento da variável resposta em relação a todas as variáveis explicativas. Para termos uma visão da força de correlação entre as variáveis explicativas e a diversidade funcional, é interessante fazer uma análise de regressão múltipla passo a passo. Vamos começar com o modelo completo e dar uma olhada nos valores de r^2 , r^2 ajustado e de p .

Tabela 21 – Resultado do modelo completo de regressão múltipla

Efeito	Estimativa	Erro padrão	Valor de t	Valor de p
Intercepto	2,032	0,664	3,057	0,028
TURB	-0,012	0,005	-2,38	0,063
ALCA	0,307	0,232	1,318	0,245
OD	-0,148	0,046	-3,187	0,024
COND	-0,0005	0,0002	-2,596	0,049
TEM	-0,045	0,012	-3,532	0,017
PH	0,026	0,043	0,594	0,579

Fonte: Elaborada pelo autor (2021).

Na Tabela 21, vemos que há 3 variáveis com valor de p significativos. O valor de p dos testes- t individuais é um dos critérios utilizados para selecionar as variáveis para um modelo final reduzido. Se utilizarmos esse critério de seleção, as variáveis OD, COND e TEM seriam as mais importantes. Porém, há duas questões importantes que devem ser levadas em conta para montar um modelo final reduzido: o número de variáveis explicativas e o tamanho amostral. Esses valores de p dos testes- t não levam em consideração essas questões. Então, temos de utilizar um método que leve em consideração esses detalhes importantes.

Vamos analisar, agora, os métodos de seleção de variáveis mais adequados. O resultado para o teste de seleção para frente ajustada foi um modelo final reduzido com somente a variável ALCA com $r^2 = 0,612$, r^2 ajustado = 0,573 e valor de $p = 0,002$. O resultado do teste AIC foi igual, com a seleção somente da variável ALCA e um AIC de -57,39, comparado ao AIC do modelo completo de -48. Na análise com o AIC, quanto menor o valor, melhor o modelo. Então, a diversidade funcional de peixes nesses riachos responde fortemente à variação da variável alcalinidade, com mais diversidade em riachos mais alcalinos. Isso quer dizer que pode haver perda de diversidade se houver alteração química nos riachos.

Os resultados de uma regressão linear múltipla são apresentados com a tabela e graficamente. No gráfico, dependerá de quantas variáveis permanecem no modelo. Se somente uma variável permanecer, como o modelo reduzido final acima, pode-se apresentar um gráfico XY com as duas variáveis exatamente como a regressão linear simples. Se duas variáveis permanecerem, pode-se montar um gráfico tridimensional, ou, alternativamente, apresentar dois gráficos XY com cada variável explicativa. O gráfico tridimensional é analiticamente mais “correto”, pois apresenta a variação da variável resposta em relação às duas variáveis explicativas ao mesmo tempo. Porém, nem sempre é fácil de visualizar.

6.3.8 Análise de covariância

Análise de covariância (ANCOVA) combina as técnicas de análise de variância (ANOVA) e regressão linear simples com a incorporação de uma variável quantitativa (covariável) junto com a variável nominal em um único modelo. O uso principal da ANCOVA é aumentar a precisão de experimentos aleatorizados (em que os objetos foram designados aleatoriamente aos tratamentos). Uma covariável X é medida em cada unidade amostral antes do tratamento ser aplicado. As médias de cada grupo (tratamento) são ajustadas para remover as diferenças iniciais, reduzindo, portanto, o erro experimental e permitindo uma comparação

mais precisa entre os grupos. Outro uso é modelar relações quando se quer comparar regressões em grupos de uma variável nominal qualquer. O pressuposto para esse uso é diferente do que o uso anterior. O último uso da ANCOVA é retirar (controlar) fontes de variação em experimentos naturais/observacionais.

Há duas hipóteses testadas na ANCOVA. A primeira é que as inclinações das retas das regressões lineares dos dois grupos são iguais, ou seja, que as duas retas são paralelas. A segunda hipótese é que os interceptos são iguais, ou seja, que as médias ajustadas são iguais. Então, o primeiro passo em uma ANCOVA é computar as retas de regressão linear com a covariável para cada grupo separadamente. As inclinações dessas retas devem ser iguais. Em outras palavras, elas devem ser paralelas. Considerando que as retas são paralelas e tem a mesma inclinação, o segundo passo da ANCOVA é comparar os interceptos, que são as médias ajustadas pelas inclinações, ou seja, é a comparação das médias entre os grupos.

Durante a análise, o pesquisador deve, primeiramente, verificar a interação entre a variável nominal e a covariável (já estudamos sobre interação entre variáveis explicativas na parte sobre ANOVA de duas vias). Se houver interação, é recomendado que não se prossiga com a análise do efeito de diferença entre as médias dos grupos. Por quê? Porque a relação entre a variável resposta e a covariável não é a mesma para cada grupo. A lógica é que há um ruído que corresponde a uma variação compartilhada entre duas variáveis explicativas. No caso da ANCOVA, são a variável nominal e a covariável. Mesmo com essa dificuldade, pode ser interessante analisar o comportamento da variável resposta em cada grupo, se isso fizer sentido a partir do objetivo do estudo. A dica, para usar adequadamente uma ANCOVA, é que o objetivo seja bem claro antes de realizar o experimento, pois isso dará a base biológica para ele. Dessa forma, qualquer comportamento da variável resposta pode fazer sentido, o que é mais importante. O que não se pode fazer é usar a covariável em uma ANCOVA e descartá-la quando perceber que ela influencia a comparação da média da variável resposta dos grupos, principalmente se uma ANOVA se torna significativa sem a covariável no modelo. É mais honesto contar toda a história da análise.

Os pressupostos da ANCOVA são os mesmos da regressão linear simples e da ANOVA. Além disso, há dois pressupostos únicos da ANCOVA. O primeiro, é que a covariável é independente do efeito entre grupos. Em outras palavras, as médias da covariável para cada grupo devem ser iguais. O segundo, é que as duas retas de regressão linear da relação entre a covariável e a variável resposta são paralelas.

6.3.8.1 Exemplo de análise de covariância

Em 2006 e 2007, experimentos de campo foram conduzidos para determinar o efeito de interferência da planta *Salvia reflexa* sobre a colheita de beterraba nos EUA (ODERO *et al.*, 2010). *S. reflexa* foi plantada ao lado dos pés de beterraba em densidades constantes e as plantas competiram por períodos (36 a 133 dias) após a emergência das folhas de beterraba. As colheitas (megagramas por hectares) foram determinadas no final da estação. O objetivo é analisar a influência da presença de *S. reflexa* sobre a colheita da beterraba nos dois anos. Então, temos uma variável nominal (ano) e uma variável quantitativa discreta (duração da competição) como variáveis explicativas, e uma variável quantitativa contínua (colheita) como variável resposta.

O primeiro passo é verificar se há diferenças nas inclinações das retas de regressão entre os dois anos. Para isso, realizaremos uma análise com todas as variáveis e testaremos a interação entre a duração da competição e os anos. Se houver uma diferença significativa, isso significa que o efeito da covariável é dependente do ano. Se esse for o caso, o indicado é se ater somente às variações dentro de cada ano, abandonando a comparação.

Tabela 22 – Resultado da análise de covariância com a covariável no modelo

Efeito	Soma de quadrados	Graus de liberdade	Quadrados médios	F	p
Duração	4530	1	4530	74,11	<0,001
Ano	1150	1	1150	18,82	<0,001
Duração × ano	2	1	2	0,033	0,856
Total	3667	60	61		

Fonte: Odero *et al.* (2010).

A Tabela 22 mostra que não há interação entre os efeitos da duração da competição e o ano ($p = 0,856$). Podemos seguir para a próxima etapa, que é a comparação dos interceptos (médias dos grupos). Para isso, reanalisamos os dados retirando o efeito da interação, tendo em vista que ele não é significativo.

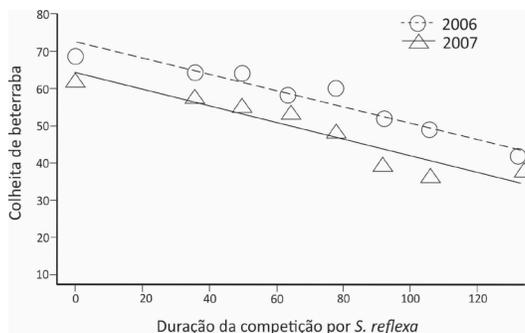
Tabela 23 – Resultado da análise de covariância sem a covariável

Efeito	Soma de quadrados	Graus de liberdade	Quadrados médios	F	p
Duração	4530	1	4530	75,31	<0,001
Ano	1150	1	1150	19,12	<0,001
Total	3669	61	60		

Fonte: Odera *et al.* (2010).

A Tabela 23 mostra que há diferença nas médias da quantidade de colheita entre os dois anos e, também, que houve uma influência significativa da duração da competição por *S. reflexa* sobre a colheita. Como não houve interação significativa entre as duas variáveis explicativas, a variação na colheita entre os dois anos não está relacionada com a intensidade de competição por *S. reflexa*. Ou seja, houve resposta da colheita à competição, mas isso não influenciou diretamente na diferença de colheita entre os dois anos. Podemos olhar para o padrão de variação da colheita em relação à competição nos dois anos para ter uma ideia do que ocorreu.

Figura 24 – Relação entre duração de competição e colheita de beterraba em dois anos



Fonte: Modificado de Odera *et al.* (2010).

A Figura 24 mostra a relação entre a duração da competição por *S. reflexa* e a colheita de beterraba para os anos de 2006 e 2007. Pode-se notar que a influência da competição foi negativa nos dois anos. Quanto maior a densidade de *S. reflexa*, menor a colheita de beterraba. Houve uma diferença na média da colheita entre os anos, mostrada na diferença entre os dois interceptos, mas

essa diferença não esteve associada com a intensidade de competição, como vimos nos resultados da ANCOVA. O intervalo de confiança para a variável explicativa ano foi -12,36 a -4,60, indicando que se deve abandonar a hipótese nula de que a diferença das médias é zero. O intervalo de confiança para o coeficiente β_1 da variável explicativa duração da competição foi -0,26 a -0,17.

6.3.9 Análise de regressão logística

Até aqui, estudamos testes estatísticos que assumem que a variável aleatória de interesse é normalmente distribuída e que as variâncias são homogêneas. Esses testes são conhecidos genericamente por modelos lineares gerais (do inglês *general linear models*). No caso de a variável resposta não ser normalmente distribuída, vimos que uma opção, para usar os testes clássicos, é transformá-la. Mas, por que não podemos usar a variável com sua própria distribuição? Bem, vimos que testes de permutação podem ser usados para fazer isso, já que não exigem o pressuposto de normalidade. Na verdade, para esses testes a variável aleatória de interesse pode assumir qualquer distribuição (MANLY, 2006). Porém, assim como podemos modelar uma resposta da variável aleatória com base na distribuição normal, também é possível modelar com base em outras distribuições. Isso nos leva a um conjunto de métodos conhecido como modelos lineares generalizados (do inglês *generalized linear models*) (MCCULLAGH; NELDER, 1989). Um modelo linear generalizado é uma generalização flexível dos modelos lineares clássicos, que permite que a variável resposta possua distribuição diferente da normal. Não está dentro do escopo do livro apresentar os modelos lineares generalizados. Porém, para finalizar este capítulo sobre os testes estatísticos, estudaremos um teste que é utilizado com frequência com dados biológicos.

Uma aplicação importante dos modelos lineares generalizados para dados biológicos é a modelagem de uma variável resposta que é binária (presença/ausência ou vivo/morto por exemplo). Quando há uma variável quantitativa contínua como variável explicativa, o teste estatístico utilizado é a regressão logística simples. No caso da variável resposta binária, realmente se modela a probabilidade $Y = \Pi(x)$. Ou seja, a probabilidade de Y tomar um dos valores de X . Neste caso, a distribuição não é mais normal, mas binomial, pois há apenas dois valores possíveis da variável Y . Portanto, não se pode usar o método dos quadrados mínimos, como na regressão linear e as estimativas dos coeficientes do modelo devem ser calculadas com o método chamado de máxima verossimilhança. Há dois objetivos na análise: 1 – saber se a probabilidade de encontrar um valor da

variável nominal está associada com a variável quantitativa; e 2 – prever a probabilidade de encontrar um valor particular da variável nominal dado um valor da variável quantitativa. A hipótese nula da regressão logística é que a probabilidade de ocorrência de um valor particular da variável nominal não está associada com a variável quantitativa, ou, em outras palavras, que a linha que descreve a relação entre as duas variáveis tem uma inclinação igual a zero.

O teste encontra a equação que melhor prevê o valor da variável Y para cada valor da variável X . A diferença para a regressão linear é que não se mede diretamente a variável Y , pois só há dois valores possíveis. Na verdade, calcula-se a probabilidade de se obter um valor particular da variável nominal para cada valor de X . A variável Y usada será, então, a probabilidade de o evento que está sendo medido ocorrer ao longo da amplitude de valores da variável X . Essa probabilidade varia de 0 a 1. Na análise, se utiliza a probabilidade de que $Y = 1$ relativamente à probabilidade de que $Y = 0$ (QUINN; KEOUGH, 2002):

$$\frac{\Pi(x)}{1 - \Pi(x)}$$

Se as chances de ocorrência > 1 , então a probabilidade de que $Y = 1$ é maior do que a probabilidade de que $Y = 0$; se as chances são < 1 , então o inverso é verdadeiro. Depois, se toma o logaritmo natural:

$$\ln \left[\frac{\Pi(x)}{1 - \Pi(x)} \right]$$

Nomeamos essa probabilidade como $g(x)$, que pode ser modelada em relação a uma variável explicativa com uma forma que lembra o modelo linear tradicional:

$$g(x) = \beta_0 + \beta_1 X$$

Encontra-se a inclinação β_1 e o intercepto β_0 do melhor ajuste da equação utilizando um método chamado de máxima verossimilhança. Essa técnica de máxima verossimilhança encontra os valores dos parâmetros sob os quais se teria a melhor probabilidade de encontrar os valores observados. O valor de p pode ser encontrado com técnicas diferentes. Porém, recomenda-se o uso da razão de verossimilhança, a qual usa a diferença entre a probabilidade de obter os resultados encontrados sob o modelo logístico e a probabilidade de obter os resultados encontrados se não houver relação entre as duas variáveis (modelo nulo). A regressão logística assume que as observações são independentes, e

que a relação entre o logaritmo natural da razão de chance e a variável quantitativa é linear. É possível calcular os intervalos de confiança para o intercepto e para a inclinação da curva.

6.3.9.1 Exemplo de regressão logística

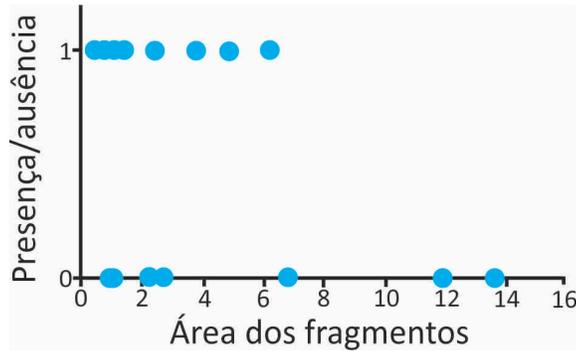
Analisaremos a probabilidade de ocorrência da aranha *Tidarrem haemorrhoidale* em 16 fragmentos florestais da mata Atlântica do sul do Brasil. A variável resposta binária é presença/ausência da espécie e a variável quantitativa é área de cada fragmento. Então, um objetivo é determinar se há uma relação entre a presença da espécie e a área dos fragmentos florestais, e o outro, é predizer a probabilidade de a espécie estar presente em fragmentos com determinados tamanhos. A Tabela 24 mostra a presença e ausência da espécie e a área dos fragmentos, e a Figura 25 a distribuição da variável resposta em relação à variável explicativa.

Tabela 24 – Ocorrência de uma espécie de aranha em fragmentos florestais com diferentes áreas

Área (ha)	Presença/ausência
3,798507	1
2,373975	0
11,99711	0
6,864291	0
4,901955	1
0,830087	1
0,363559	1
6,260152	1
13,69452	0
0,606617	1
2,759371	0
1,069333	0
1,259256	1
2,395813	1
1,286634	0
1,053435	1

Fonte: Baldissera (2012).

Figura 25 – Presença/ausência de uma espécie de aranha em 16 fragmentos florestais



Fonte: Elaborada pelo autor (2021).

A Tabela 25 mostra os resultados do teste. O valor de p é fornecido automaticamente pelo programa e não é o recomendável.

Tabela 25 – Resultado da regressão logística

Efeito	Estimativa	Erro padrão	Valor de z	p
Intercepto	1,2298	0,8153	1,508	0,131
Área	-0,2654	0,1812	-1,465	0,143

Fonte: Elaborada pelo autor (2021).

O valor de p maior do que 0,05 faz com que aceitemos a hipótese nula de que não há uma relação entre a presença dessa espécie de aranha e as áreas dos fragmentos. O valor de p baseado no teste de razão de verossimilhança é 0,08. Ou seja, ainda aceitamos a hipótese nula. Com as informações da tabela, montamos a equação:

$$Y = e^{1,2298 - 0,2654(\text{Área})} / (1 + e^{1,2298 - 0,2654(\text{Área})})$$

Nesta equação, e é a raiz do logaritmo natural s . Então, nesse caso, para prever a probabilidade de ocorrência da espécie em um fragmento específico, pode-se medir a área do fragmento, colocar o valor na equação (Área) e conseguir uma estimativa de Y , ou a probabilidade de a espécie estar naquele fragmento. Assim como a regressão linear, a regressão logística também pode ser realizada com mais de uma variável quantitativa, chamando-se regressão logística múltipla. O intervalo de confiança para o coeficiente β_0 é -0,25 a 3,07 e de β_1 é -0,73 a 0,03.

7.1 PROBABILIDADE

Apesar de a estimativa de probabilidades ser uma parte fundamental da estatística, raramente será preciso fazer os cálculos. Mas é prático saber algumas regras básicas sobre adição e multiplicação de probabilidades.

A ideia básica de um teste estatístico é identificar uma hipótese nula, coletar alguns dados e, então, estimar a probabilidade de se conseguir os dados observados se a hipótese nula é verdade. Se a probabilidade de se conseguir um resultado igual ao observado é baixa, considerando a hipótese nula, se conclui que a hipótese nula provavelmente não é verdadeira. É útil, então, saber um pouco sobre probabilidade.

Uma maneira de pensar sobre probabilidade é como a proporção de indivíduos de uma população que possui uma característica particular. A probabilidade de amostrar um determinado tipo de indivíduo é igual à proporção desse tipo na população. Por exemplo, em uma universidade havia 8008 matriculados em 2019, dos quais 274 eram pós-graduandos. A probabilidade de você amostrar aleatoriamente um estudante que fosse pós-graduando era $274/8008$, ou 0,034216. Em outras palavras, 3,421578% ($0,034126 \times 100$) dos estudantes eram pós-graduandos, então se você escolher um estudante aleatoriamente, a chance de que seja de pós-graduação é de em torno de 3,42%.

Quando se lida com probabilidades na área biológica, pode-se trabalhar com expectativas teóricas, não amostras de populações. Por exemplo, em um acasalamento genético de dois indivíduos de mosca que são heterozigotos no alelo *vestigial*, a teoria de Mendel prevê que a probabilidade de um indivíduo da prole ser homozigoto (asas vestigiais) recessivo é de um quarto, ou 0,25. É o mesmo que dizer que um quarto da população da prole terá asas vestigiais.

Você calcula a probabilidade de um indivíduo ter um valor de uma variável nominal e outro valor de uma segunda variável nominal pela multiplicação das probabilidades de cada valor juntos. Por exemplo, na F2 dos cruzamentos do experimento de Mendel, um indivíduo Aa produz gametas A e a na proporção de 50% para cada um. Para formar um indivíduo AA, precisa ocorrer o encontro simultâneo de dois gametas A. A probabilidade de este evento ocorrer é a mul-

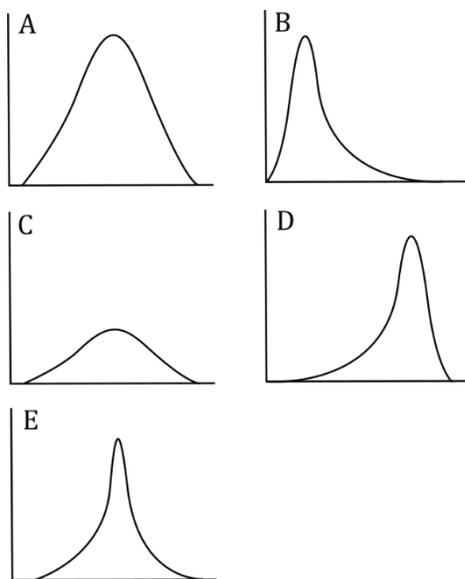
tiplicação dos eventos isolados: $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$. A mesma fórmula pode ser usada para um descendente aa.

A probabilidade de que um indivíduo tenha um valor **OU** outro valor, mutuamente exclusivo, é encontrada pela adição das probabilidades de cada um deles. Mutuamente exclusivo significa que um indivíduo não pode ter os dois valores ao mesmo tempo. Por exemplo, se a probabilidade de uma flor em um cruzamento genético ser vermelha é um quarto, então a probabilidade de que seja vermelha **OU** rosa é um quarto mais um meio ou três quartos.

7.2 ASSIMETRIA E CURTOSE

Como já estudamos, uma forma de analisar a distribuição de frequências de um conjunto de dados é por meio de um histograma. O histograma mostrará, para uma variável normalmente distribuída, a característica forma de sino (Figura 26-A). Com uma cauda longa no lado direito, a distribuição é assimétrica para a direita (B). Um histograma com uma cauda longa à esquerda é chamado de assimétrico para a esquerda (D). Há uma estatística para descrever a assimetria, g_1 , mas não há nenhuma razão para calculá-la. Não há uma regra que não se deva realizar um teste clássico se g_1 é maior do que algum valor de corte.

Figura 26 – Histogramas de frequências



Fonte: Elaborada pelo autor (2021).

Outra forma em que os dados podem se desviar da distribuição normal é a curtose. Um histograma com um meio largo e plano e caudas curtas é platicúrtico (Figura 26-C). Um histograma que tem um pico alto no meio e longas caudas nos dois lados é chamado de leptocúrtico (E). A estatística para descrever a curtose é g_2 , mas aparentemente não há razão para cálculo também.

7.3 TESTE DE PERMUTAÇÃO

Muitas hipóteses de interesse na ciência podem ser vistas como alternativas à hipótese nula de aleatoriedade. Isso quer dizer que a hipótese de estudo sob investigação sugere que haverá uma tendência de um padrão aparecer nos dados, enquanto a hipótese nula afirma que, se esse padrão estiver presente, então é puramente um efeito das observações ordenadas aleatoriamente.

O teste de permutação é uma forma de determinar se a hipótese nula é razoável nesse tipo de situação (MANLY, 2006). Uma estatística E (diferença entre médias, soma de quadrados, coeficiente de correlação, etc.) é escolhida para medir a extensão na qual os dados mostram o padrão observado. Um valor e de E para os dados observados é então comparado com a distribuição de E obtida pelo reordenamento aleatório dos dados. O argumento aqui é que se a hipótese nula é verdadeira, então todas as ordenações possíveis dos dados possuem a mesma probabilidade de ocorrer. A ordenação dos dados observados é, então, somente uma das igualmente prováveis outras possibilidades e e deve ocorrer como um valor típico da distribuição aleatória de E . Se esse não parece ser o caso (de forma que e é significativa), então a hipótese nula é descartada até certo ponto e, por consequência, a hipótese alternativa é considerada mais plausível.

O nível de significância de e é a proporção ou percentual de valores que são iguais ou mais extremos do que esse valor e na distribuição aleatória. A interpretação é a mesma usada nos testes de significância convencionais. Se a frequência de ocorrências de e é menor do que 5%, então há evidência de que a hipótese nula não é verdadeira; se a frequência de ocorrências de e é menor do que 1%, então há forte evidência de que a hipótese nula não é verdadeira. Sempre é bom olhar o nível de significância como uma medida da força de evidência contra a hipótese nula, ao invés de se os dados são significativos em um certo nível.

Dessa forma, esse tipo de teste estatístico não necessita cumprir com alguns pressupostos dos testes clássicos, o que facilita o desenho, coleta e análise. Porém, a independência das observações deve ser cumprida e a homocedastici-

dade é desejável. A Tabela 26 mostra 22 medidas do comprimento do corpo de machos e fêmeas de aranhas da espécie *Dipoena granulata* (em mm) coletadas em uma unidade de conservação da Mata Atlântica (GALETI, 2020).

Tabela 26 – Comprimentos de corpos de aranhas machos e fêmeas

Machos	1,9	1,9	1,7	1,6	1,6	1,9	1,8	1,8	1,5	1,6	1,6
Fêmeas	1,6	1,6	2	1,3	1,9	1,7	1,5	1,6	1,8	1,9	1,5

Fonte: Galeti (2020).

As médias arredondadas para os dados observados foram:

Média dos machos = $1,9+1,9+1,7+1,6+1,6+1,9+1,8+1,8+1,5+1,6+1,6/11 = 1,718$ mm

Média das fêmeas = $1,6+1,6+2+1,3+1,9+1,7+1,5+1,6+1,8+1,9+1,5/11 = 1,673$ mm

E a diferença foi de $1,718 - 1,673 = 0,045$ mm.

A ideia básica da permutação é:

1. Calcular o tamanho do efeito. Nesse caso, estamos interessados em calcular a diferença entre as médias do comprimento corporal nos dois sexos, que é = 0,045 mm;
2. Permutar as posições dos valores de cada objeto;
3. Recalcular a estatística, ou seja, a diferença entre as médias;
4. Repetir n vezes (1000, 10000) os passos 2 e 3;
5. Montar a distribuição das frequências dos valores das n diferenças entre as médias;
6. Contar quantas vezes na distribuição de frequências houve um valor de diferença entre as médias mais extremo do que o observado em 1;
7. Dividir o número de vezes que a diferença foi mais extrema pelo número total de permutas. Esse é o valor de p .

Por exemplo, uma possível permuta para os dados acima é:

Tabela 27 – Comprimentos de corpos de aranhas machos e fêmeas permutados

Machos	1,6	1,9	1,7	1,6	1,6	1,9	1,8	1,8	1,5	1,6	1,6
Fêmeas	1,9	1,6	2	1,3	1,9	1,7	1,5	1,6	1,8	1,9	1,5

Fonte: Galeti (2020).

As médias para a segunda combinação foram machos=1,69 e fêmeas=1,7, com uma diferença de -0,01. O valor de p é calculado pela probabilidade de se encontrar uma diferença igual ou mais extrema à observada (0,045 mm) entre todas as diferenças calculadas para todos as combinações possíveis (nesse caso, há 705432 combinações possíveis!), ou para um número de combinações determinado (por exemplo, 1000). Para os dados observados, o valor do teste- t clássico, com 20 graus de liberdade, é 0,588 com um valor de $p = 0,563$. Ou seja, a probabilidade de haver uma diferença entre as duas médias sob a hipótese nula é baixa. O valor de p para o teste de permutação foi de 0,654, muito parecido com o teste clássico. Isso significa que a diferença entre as médias observada (0,045) ocorreu com muita frequência na distribuição e, portanto, essa diferença não é rara de ocorrer sob a hipótese nula. Note que, para o caso do teste clássico, que é baseado em uma distribuição de frequências teóricas, o valor de p será sempre o mesmo, em qualquer programa. Porém, para o teste de permutação, como, em geral, se realiza menos combinações do que as possíveis, o valor de p de uma nova análise pode ser ligeiramente diferente, pois será resultado de outras combinações.

Deve-se ter em mente que muitas áreas das ciências biológicas ainda se utilizam de testes clássicos e a publicação de testes heterodoxos pode ser vista com cautela por editores de periódicos científicos e pelos pares do pesquisador. Assim, sempre que se optar por utilizar uma ou outra abordagem na pesquisa, deve ficar bem claro porque a abordagem foi escolhida e, caso a abordagem clássica seja usada, se os dados cumprem com os pressupostos mínimos.

7.4 TESTES MÚLTIPLOS

Quando vários testes de hipóteses são realizados simultaneamente no mesmo conjunto de dados, a probabilidade de ocorrer um erro Tipo I fica maior do que o valor nominal de significância estabelecido (normalmente 5%, ou um $\alpha = 0,05$). Por exemplo, em uma matriz de correlação entre 5 variáveis, 10 testes são realizados simultaneamente. Para dados gerados aleatoriamente, há uma probabilidade $p = 0,401$ (computado da distribuição binomial) de rejeitar a hipótese nula pelo menos uma vez em 10 testes com um p nominal de 0,05. Ou seja, é muito mais provável rejeitar pelo menos uma hipótese nula (41% das vezes) quando se compara com o valor de p nominal 0,05 (5% das vezes). Esse erro é chamado de taxa de erro em relação à família (termo em português livre-

mente traduzido de *familywise*). Então, quando se conduzem múltiplos testes, deve-se realizar um teste global da significância para determinar se há, mesmo, algum valor significativo no conjunto de variáveis.

Uma abordagem geral é a correção de Bonferroni (1935) para k testes independentes. Nesse caso, substitui-se o nível de significância, digamos $\alpha = 0,05$ por um nível ajustado $\alpha' = \alpha/k$ e se comparam as probabilidades resultantes do teste com α' . O teste de Bonferroni é considerado muito conservador (NARUM, 2006), o que significa que ele leva a rejeitar poucas vezes a hipótese nula quando a hipótese alternativa é verdadeira, o que diminui o poder do teste. Mas, é o teste de correção mais usado e conhecido.

Há dois tipos de testes múltiplos: dependentes e independentes. Os testes dependentes são aqueles nos quais partes dos dados são compartilhadas pelos diferentes testes, enquanto nos independentes, cada teste usa seus próprios dados. Legendre e Legendre (2012) sugerem, para o primeiro caso, o uso da correção de Holm, enquanto, para o segundo caso, sugerem a correção de Hochberg. De qualquer forma, sempre que as análises envolverem mais de um teste de hipótese, é importante que o pesquisador esteja consciente de que deve utilizar uma correção no valor nominal de p , sob pena de aumentar a chance de cometer um erro Tipo I.

7.5 TESTES NÃO PARAMÉTRICOS CLÁSSICOS

7.5.1 Mann-Whitney

O teste U de Mann-Whitney se baseia na ordenação dos valores da variável resposta. Ele funciona muito bem com variáveis ordinais. Em geral, ele é sugerido como alternativa ao teste- t para duas amostras de dados não normais. Entretanto, se as duas amostras possuem a mesma distribuição (pode ser não normal!), o teste- t não será sensível a desvios da normalidade e, portanto, pode-se usá-lo sem problemas. Se as duas amostras possuem distribuições diferentes, o teste U de Mann-Whitney não é melhor do que o teste- t . Então, não há razão para usar o teste U de Mann-Whitney, a menos que se tenha realmente uma variável ordinal ao invés de uma variável quantitativa.

Se os desvios-padrões são muito diferentes (um desvio-padrão é duas ou mais vezes maior do que o outro) e os tamanhos amostrais são pequenos (me-

nos que 10) ou diferentes, deve-se usar o teste- t de Welch. Esse teste é similar ao teste- t de Student, mas não assume que os desvios-padrões são iguais. Esse teste pode ser muito mais acurado quando os desvios-padrões são muito diferentes. Dessa forma, a melhor utilização do teste U de Mann-Whitney é quando queremos comparar a resposta de uma variável ordinal entre dois grupos.

7.5.2 Kruskal-Wallis

Esse teste, assim como o anterior, também é baseado em ordenações dos valores da variável resposta. Portanto, ele não possui o pressuposto de normalidade. O uso mais comum desse teste é em substituição de uma ANOVA quando os dados não se adequam ao pressuposto da normalidade. Entretanto, uma ANOVA não é muito sensível a desvios da normalidade. Além disso, há a opção de utilizar testes de permutação para comparação de médias.

O outro pressuposto da ANOVA, a homocedasticidade, também é pressuposto do teste de Kruskal-Wallis. Ou seja, esse teste não é melhor do que uma ANOVA se os dados possuem heterocedasticidade. O teste correto para essa situação é a ANOVA de Welch. A hipótese do teste de Kruskal-Wallis é de que não há diferença na localização das distribuições entre os grupos analisados, o que é diferente da hipótese de uma ANOVA. Dessa forma, a melhor utilização do teste de Kruskal-Wallis é quando queremos estudar a resposta de uma variável ordinal entre vários grupos.

7.5.3 Correlação de postos de Spearman

A correlação de postos de Spearman é utilizada quando se tem duas variáveis e se quer saber se as duas estão associadas; se, na medida em que uma variável aumenta, a outra variável tende a aumentar ou diminuir. Também se utiliza esse tipo de correlação se há uma variável quantitativa e uma variável ordenada. Nesse caso, converte-se a variável quantitativa em postos e se usa a correlação de Spearman nos dois conjuntos de ordenamentos.

Algumas pessoas usam a correlação de Spearman como uma alternativa não paramétrica à regressão linear ou correlação quando possuem duas variáveis quantitativas e uma, ou as duas não possuem distribuição normal. Isso requer que se converta as duas variáveis para postos (ordenação). Porém, muitos estudos de simulação mostraram que a regressão linear e a correlação não são

sensíveis à falta de normalidade. Uma ou as duas variáveis podem ser bem não normais e a probabilidade de um falso positivo é ainda em torno de 0,05 (EDGELL; NOON, 1984). Não é incorreto usar a correlação de Spearman para duas variáveis quantitativas, mas a regressão linear e a correlação são muito mais comumente usadas e são familiares a muito mais pessoas. Então, é recomendável usar a regressão linear e a correlação sempre que você tiver duas variáveis quantitativas, mesmo que elas não pareçam normais. Se for o caso, pode-se usar correlação e regressão por permutação também.

Quando você usa a correlação de Spearman, não há nenhum pressuposto de normalidade ou homocedasticidade. Também não se assume que a relação é linear. Você pode usar a correlação de Spearman mesmo que a associação entre as variáveis seja curva, desde que a relação seja monotônica (quando X fica maior, Y continua ficando maior, ou menor).

A correlação de Spearman calcula o valor de p da mesma maneira que a regressão linear e a correlação, exceto que se faz nos postos, não nas medidas. Para converter uma variável quantitativa em postos, o maior valor se torna 1, o segundo maior se torna 2 e, assim, sucessivamente. Se há valores iguais, usa-se a média dos postos. Por exemplo, se duas observações são iguais para o segundo maior posto, então o valor do posto é 2,5 (a média de 2 e 3).

Na correlação de Spearman, o coeficiente de correlação de Pearson (r) é, agora, o coeficiente de correlação de Spearman (ρ) e se pode usá-lo como uma medida da força de associação. Quase nunca se usará uma linha de regressão para descrever ou prever quando se usa a correlação de Spearman, então, não é necessário calcular o equivalente de uma linha de regressão.

7.6 ANÁLISE DE PODER

O poder estatístico de um teste de hipótese é a probabilidade de detectar um efeito se esse efeito realmente é verdadeiro. A análise de poder pode ser realizada para estimar o número de objetos (tamanho amostral) necessário para detectar um efeito em um experimento. Isso é especialmente verdade se o pesquisador estiver pretendendo realizar algum experimento que envolva dor em humanos ou outros vertebrados, quando é particularmente importante minimizar o número de indivíduos (sem tornar o tamanho amostral tão pequeno que todo experimento não passe de perda de tempo e sofrimento), ou se um experimento que demandará muito tempo ou muito dinheiro for planejado. Métodos

foram desenvolvidos para muitos testes estatísticos a fim de estimar o tamanho amostral necessário para detectar um efeito particular, ou para estimar o tamanho de um efeito que pode ser detectado com um tamanho amostral particular. Para realizar uma análise de poder, é necessário especificar, basicamente, quatro parâmetros: tamanho de efeito, poder estatístico desejado, desvio-padrão e nível de significância.

O tamanho do efeito se refere ao tamanho ou magnitude de um efeito ou resultado que é esperado de ocorrer na população. Por exemplo, se o objetivo do estudo é comparar duas médias, o tamanho do efeito é a diferença esperada entre essas duas médias. Esse tipo de tamanho de efeito não é padronizado, uma vez que diferentes estudos com os mesmos objetos podem estimar uma diferença diferente entre as médias. Há outras formas de calcular o tamanho de efeito padronizado, uma delas é o valor d de Cohen (COHEN, 1988). Para pesquisas biológicas aplicadas e clínicas, pode haver um tamanho de efeito bem definido que se deseja detectar. Quando se faz pesquisa de biologia básica ou experimentos naturais/observacionais, em geral não se sabe o quão grande uma diferença deve ser. Nesse caso, talvez qualquer diferença seja relevante e o ideal é ter o maior número de objetos que seu experimento pode comportar. O tamanho do efeito é o desvio mínimo da hipótese nula que você espera encontrar. O tamanho do efeito pode ser mais informativo para o público em geral do que o valor de p . Um valor de p alto em um estudo com pequeno tamanho amostral pode ser encontrado, mas o tamanho do efeito pode ser relevante para propósitos práticos, principalmente se afeta o bem-estar e a saúde das pessoas. Então, o pesquisador não deve simplesmente se ater ao resultado estatístico da sua pesquisa, mas ao significado biológico encontrado. Um resultado significativo prático é significativo no mundo real (ELLIS, 2010).

O poder estatístico é a probabilidade de detectar o efeito especificado (tamanho do efeito). Em geral, estabelece-se um alto poder, tal como 0,8 (80%) ou 0,9 (90%). O custo de um falso negativo (erro Tipo II) deve influenciar a escolha do poder. Se alguém realmente quer ter certeza de que detectou o tamanho do efeito em questão, quererá usar um alto valor de poder, o que resultará em um tamanho amostral maior. O comum é usar 80%.

Para variáveis quantitativas, também será necessária uma estimativa do desvio-padrão. À medida que o desvio-padrão fica maior, torna-se mais difícil detectar uma diferença significativa, então será preciso um tamanho amostral maior. A estimativa do desvio-padrão pode vir de um experimento piloto ou de experimentos similares publicados na literatura. O desvio-padrão, depois de

fazer o experimento provavelmente não será exatamente o mesmo, então o experimento irá em verdade ser alguma coisa mais ou menos poderoso.

O nível de significância do teste é o valor de p , a probabilidade de rejeitar a hipótese nula mesmo que ela seja verdadeira (um falso positivo). O valor usual é $\alpha = 0,05$.

O exemplo a seguir mostra o passo a passo para decidir um tamanho amostral baseado na análise de poder, tomando em consideração o objetivo de comparar duas médias. A hipótese nula é que não há diferença entre as médias e a hipótese alternativa é que há diferença entre as médias (bicaudal). Os passos para calcular o tamanho amostral necessário são os seguintes:

1. Especificar o nível de significância – tradicionalmente é 5% ou 0,05.
2. Decidir qual poder, ou a chance de detectar um efeito real se ele está presente. Aqui, é importante pensar no tipo de experimento. Se as consequências de falhar em detectar o efeito (erro Tipo II) forem sérias, tipo em testes de toxicidade, se deve estabelecer um poder alto, tal como 90%. Por outro lado, em estudos básicos, ou que não envolvam maiores riscos, e que se está interessado em grandes efeitos, um poder de 80% pode ser suficiente para detectar esses efeitos e menos objetos serão necessários.
3. Obter uma estimativa de desvio-padrão. Essa estimativa deve vir de um experimento prévio, da literatura ou de um experimento piloto. Se for usar o que já foi publicado, pode ser melhor pesquisar vários estudos e fazer uma média informal do desvio-padrão. Também pode ser útil calcular dois tamanhos amostrais: um com a melhor (maior?) e outro com a pior (menor?) estimativa. Veja que maior ou menor depende da hipótese de estudo.
4. Estimar o tamanho da diferença entre as médias. Novamente, pode ser retirado de experimentos prévios, da literatura ou de um experimento piloto. De qualquer forma, essa diferença deve fazer sentido. Quanto menor a diferença entre as médias, maior o tamanho amostral que será preciso. Pode-se usar a diferença entre as médias, ou um efeito padronizado, tal como (média 1 – média 2 / desvio-padrão).

7.7 TRANSFORMAÇÕES DE DADOS

A transformação de dados é uma prática utilizada em algumas áreas, enquanto, em outras, atualmente, se utiliza um teste de permutação, por exemplo. Os testes de permutação são muito utilizados na ecologia. Reexpressão de da-

dos pode ser a forma de melhor denominar a transformação de dados. Transformação pode levar à ideia de manipulação. Como vimos ao longo do livro, os testes paramétricos podem ser muito robustos a desvios dos pressupostos. Se os dados são razoavelmente distribuídos e os desvios-padrões razoavelmente homogêneos, provavelmente não há razão para realizar uma transformação nos dados. O que não se deve fazer é procurar por uma transformação que torne os resultados significativos, isso é desonesto.

Transformar dados é fazer uma operação matemática em cada medida. Esse novo valor é, então, usado nas análises estatísticas. Há várias transformações disponíveis, mas o ideal, se realmente for preciso usar, é fazer uma que faça sentido na área de conhecimento da pesquisa. Após realizar a transformação, os pressupostos de normalidade e/ou homocedasticidade devem ser testados novamente. Aqui, mostraremos as transformações mais utilizadas nas pesquisas.

7.7.1 Logaritmo

Consiste em tirar o logaritmo de cada medida. Pode-se usar uma base 10 ou uma base natural, o que não influencia no teste. Mas deve-se indicar, no material e métodos do artigo ou relatório, qual logaritmo foi usado. Se há zeros ou números negativos, não se pode calcular o logaritmo. Então, o que se pode fazer é adicionar algum número inteiro a esses valores. Se há números discretos, em geral soma-se 0,5 ou 1 aos zeros. Muitas variáveis em biologia têm distribuição log-normais, o que faz com que, após transformá-las em logaritmos, elas fiquem normais.

7.7.2 Raiz quadrada

Consiste em tirar a raiz quadrada de cada medida. Se há números negativos, não se pode fazer essa operação e se deve adicionar uma constante a cada número. Uma observação sobre essa transformação é que números inteiros (0, 1, 2, etc.) se comportam diferente de números entre 0 e 1. Se há uma variável que possui valores inteiros e fracionários menores que 1, então, essa transformação não é aconselhável.

7.7.3 Arco seno

Consiste em calcular o arco seno da raiz quadrada de uma medida. Esses números devem estar entre 0 e 1. Essa transformação é usada geralmente em proporções, que variam de 0 a 1 e, também, com dados de percentagem.

8 ANÁLISES NO R

Nas páginas seguintes, serão apresentados *scripts* para realizar todos os testes apresentados neste livro no pacote estatístico R¹. O objetivo é bem direto: que o leitor possa organizar seus dados e saiba realizar os testes propostos. Não há nenhuma pretensão de aprofundar o aprendizado no uso do pacote estatístico R, o que seria inviável. Se o leitor seguir as instruções exatamente, deve conseguir realizar todos os testes dos exemplos e, também, realizar seus próprios testes.

É fortemente aconselhável estudar mais sobre a natureza e a lógica desse pacote estatístico poderoso e, até certo ponto, intuitivo. O R é constantemente aperfeiçoado pela comunidade de pesquisadores mundiais com atualizações de seus pacotes e a criação de novos pacotes em todas as áreas da ciência. Para uma abordagem do aprendizado da estatística, sugere-se o livro de James *et al.* (2013). Para uma introdução básica ao uso do R e para análises em geral, sugere-se o livro de Crawley (2007). Dois livros em português são o de Silva (2019) para uma introdução à análise de dados e o de Mello (2013).

Os dados para as análises a seguir são apresentados em forma de tabelas. A forma de inserir os dados no R ensinada aqui é por meio de um arquivo de texto (.txt). Então, todas as tabelas devem ser copiadas para um arquivo de texto e salvas dentro do diretório padrão do R.

8.1 INTRODUÇÃO AO R

O R é um ambiente computacional e uma linguagem de programação, que se especializou em manipulação, análise e visualização gráfica de dados. O uso do R é facilitado pela presença de pacotes estatísticos, que realizam diversas análises diferentes. Quando se instala pela primeira vez o programa, ele já vem com pacotes básicos. Dentro do próprio programa, há a opção de baixar outros

¹ Links para os programas: <https://cran.r-project.org/bin/windows/base/> e <https://rstudio.com/products/rstudio/download/>.

pacotes de interesse do usuário. A grande diferença do R para muitos outros programas é que este não tem barras de opções para selecionar a análise desejada. O usuário deve escrever uma função que diz ao programa o que ele deve fazer.

Esse tutorial é super básico e objetivo. Serão apresentados comandos específicos para inserir os dados no programa e realizar as análises apresentadas no livro. Para isso, será utilizada uma interface funcional do R, que é o RStudio. É importante ressaltar que o manual do R indica que o programa rodará no idioma do Windows instalado; porém, no RStudio, isso talvez não aconteça. Neste tutorial, serão mostradas as janelas em inglês. Entretanto, isso não será um grande problema uma vez que você aprenda os comandos básicos e como interpretar os resultados.

Um alerta inicial sobre o uso do R: todas as palavras digitadas devem seguir exatamente a caligrafia correta do comando (com os argumentos) e dos arquivos que o usuário criar para inserir no programa. Se o usuário errar qualquer letra, ponto, ponto e vírgula, etc. o R retornará um aviso de erro. O programa é sensível a maiúsculas e minúsculas também. Então, se o nome do seu arquivo for `Euposso.Fazer.tudo_noR_agora.txt`, você terá que digitar exatamente esse nome para que o R reconheça! Não use acentos, til e espaços nos nomes de arquivos, ou nos títulos de colunas e linhas. Copiar os comandos a seguir para o R também pode causar dor de cabeça. Em geral, o R não reconhece alguns caracteres nas fontes dos editores de texto, por exemplo, aspas. Os comandos mostrados nas próximas páginas estão digitados com uma fonte que o R deve reconhecer. Se você copiar um comando e der erro, o melhor, então, é digitar o comando no R. Outro aviso importante, o RStudio trabalha com divisores de milhares e casas decimais diferentes do que se está acostumado no Brasil. O separador de milhar é a vírgula e o separador de decimal é o ponto. Então, 1,73 no RStudio fica 1.73; e 1.000 fica 1,000. O usuário tem de criar o arquivo já com essas alterações.

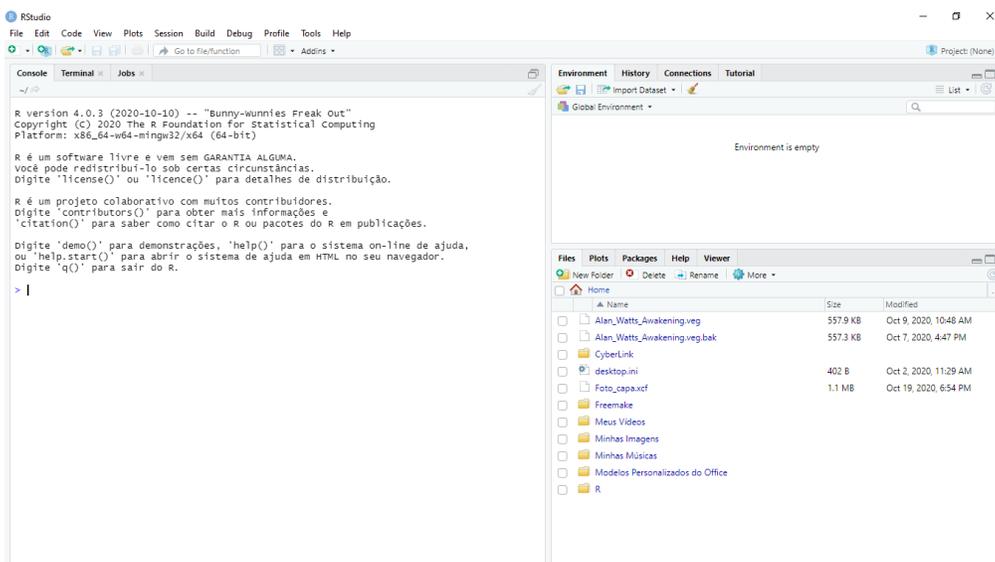
8.1.1 Área de trabalho

O pressuposto é que você já tenha instalado o R e o RStudio. Após abrir o RStudio, pode-se ver que a área de trabalho está dividida em três partes (Figura 27). Do lado esquerdo, na aba *Console*, o usuário digita os comandos desejados. Do lado direito superior, há 3 abas com opções, mas a mais usada é *History*, que armazena os comandos usados durante as sessões. Você pode chamar qualquer comando do *History* para o *Console*, simplesmente clicando duas vezes nele. Do lado direito inferior, há 5 abas de opções, incluindo a aba de *Plots*, que é onde o programa mos-

trará os gráficos que forem solicitados pelo usuário e a aba *Help*, onde o programa vai mostrar as ajudas dos comandos quando solicitado pelo usuário.

Na aba *Console*, aparecem as informações básicas sobre a versão do programa, uma declaração de falta de garantia (o programa é livre e a responsabilidade de uso é do usuário) e algumas opções de comandos para informações básicas.

Figura 27 – Área de trabalho do RStudio



Fonte: R CORE TEAM (2020).

8.1.2 Determinando o diretório de trabalho

Antes de começar a trabalhar com seus dados, você deve dizer ao programa onde eles estão localizados no computador. Salve todos os arquivos de texto das tabelas de dados na área de trabalho do Windows (ou alguma outra pasta que deseje). Agora, vá na barra de opções do RStudio em “Tools/Global options”, no campo “Default working directory (when not in a project)” procure a pasta onde guardou os arquivos com os dados. Clique em “Apply” e depois em “OK”. Se você não fizer isso, o R vai procurar qualquer arquivo que você pedir na pasta padrão dele. Então, se você trocar corretamente a pasta padrão, o programa assumirá que essa será sempre a pasta padrão.

8.1.3 Importando dados

Há várias formas de inserir os seus dados no R. Uma forma é inserir um comando para importar dados de um arquivo texto (.txt). Esse arquivo pode ser criado copiando para ele os dados da planilha do Excel ou uma tabela do Word. Nessa forma de ler os dados, o arquivo não pode ser salvo como padrão do Word (.docx) ou Excel (.xlsx). Se usar o Word ou Excel para criar o arquivo, selecione “Salvar Como” e “Texto sem formatação (.txt)”. Para evitar complicações, você pode usar o programa Bloco de Notas do Windows que salva direto em .txt. Uma tabela de dados está disponível para todas as análises que serão feitas e pode ser copiada para o arquivo texto e salva no diretório de trabalho.

8.1.4 Inserindo comandos

Logo abaixo das informações sobre o programa, há o símbolo “>”, que indica onde você deverá digitar os comandos. Por exemplo, digite, `citation()` e pressione *Enter*. O programa mostra como se deve referenciar o programa na bibliografia de seu estudo.

8.1.5 Ajuda

Para saber como funciona determinado comando, você pode digitar ? e depois o nome do comando:

```
?t.test
```

8.1.6 Calculadora

O uso mais básico do R é como calculadora. Digite os comandos a seguir e pressione *Enter* após cada um:

```
2+2
```

```
2*2
```

```
2^2
```

```
2/2
```

Note que conforme você vai digitando, os comandos ficam armazenados na aba *History*. O uso das setas ↑ e ↓ do teclado funciona diferente no R. Quando você usa essas setas, o programa chamará o histórico de comandos. Clique, agora, na seta ↑. Os comandos que você já digitou aparecem. Isso é útil para quando você quiser alterar alguma coisa do comando sem precisar digitar tudo de novo!

8.1.7 Armazenando dados

Você pode armazenar os resultados das análises em um arquivo do R:

```
soma<-2+2  
soma
```

A seta <- é formada digitando o sinal de “<” e o “-“. No comando, o R vai “criar” o arquivo “soma” e “guardar” nele o resultado da adição. Você pode criar um arquivo com números em cadeia, como uma variável, com o comando “c” (concatenar):

```
a<-c(2,2,3,5,6,7)  
a
```

A seta <- é importante para armazenar os resultados dos testes estatísticos. Sempre que você precisar saber os resultados, é só digitar o nome dado ao arquivo. Mais adiante, veremos exemplos.

8.2 HISTOGRAMA

Sintaxe

hist(x), onde “x” é uma variável quantitativa.

Dados

Cole a tabela no editor de texto, nomeie como norm.txt e grave no diretório padrão.

UMA INICIAÇÃO AOS TESTES ESTATÍSTICOS PARA DADOS BIOLÓGICOS

CLASSE	ALTURA
A	175.26
A	177.8
A	167.64
A	160.02
A	172.72
A	177.8
A	175.26
A	170.18
A	157.48
A	160.02
A	193.04
A	149.86
A	157.48
A	157.48
A	190.5
A	157.48
A	182.88
B	172.72
B	157.48
B	170.18
B	172.72
B	175.26
B	170.18
B	154.94
B	149.86
B	157.48
B	154.94
B	175.26
B	167.64
B	157.48
B	157.48
B	154.94
B	177.8
B	160.02

```

#Importando os dados para o R. Comando read.table
#O argumento h=T (header = TRUE) diz ao R que ele deve ler os títulos
das colunas
norm<-read.table("norm.txt ", h=T)

#Visualizando os dados gravados
norm

#Histograma de todos os dados
#O argumento $ diz ao R que queremos usar a variável ALTURA do arqui-
vo norm.
hist(norm$ALTURA)

#Criando um arquivo com os dados da CLASSE A
#O argumento [1:17,] diz ao R que deve pegar os dados da linha 1 a 17, e
a vírgula sem números depois expressa que são todas as colunas.
classe.A<-norm[1:17,]
classe.A

#Histograma de classe.A
hist(classe.A$ALTURA)

#Alterando o número de classes
# O argumento breaks diz ao R quantas classes usar
hist(classe.A$ALTURA, breaks = 3)

```

Desafio

1. Crie um objeto no R com os dados da CLASSE B e faça o histograma.
2. Interprete se as distribuições são normais.

8.3 TESTE DE NORMALIDADE DE SHAPIRO-WILK

Sintaxe

shapiro.test(x), onde "x" é uma variável quantitativa.

Análise

```
#Teste de normalidade da variável ALTURA da CLASSE A  
shapiro.test(classe.A$ALTURA)
```

O teste de Shapiro-Wilk tem a hipótese nula de que a distribuição de frequências da variável é igual à de uma variável com distribuição normal. O resultado mostra o valor do teste W e o valor de p .

Desafio

1. Teste a normalidade da variável da CLASSE B.

8.4 TESTE DE HOMOCEASTICIDADE DE BARTLETT

Sintaxe

`bartlett.test(y ~ x, dados)`, onde “y” é uma variável quantitativa, “x” é uma variável nominal com no mínimo duas categorias e “dados” é o nome de seu arquivo no R.

Análise

#Teste de homocedasticidade de Bartlett com os mesmos dados do teste de normalidade

```
bartlett.test(ALTURA ~ CLASSE, norm)
```

O teste de Bartlett tem a hipótese nula de que a variável nas duas classes possui a mesma variância. O resultado mostra o valor do teste $K\text{-squared}$ e o valor de p .

8.5 TESTE EXATO DO QUI-QUADRADO

Sintaxe

`chisq.test(x,p)`, onde “x” = um vetor numérico e “p” = um vetor de probabilidades do mesmo comprimento de “x”.

Dados

Copie a tabela no editor de textos, nomeie o arquivo como `qui.ex.txt` e grave no diretório padrão.

<i>Andropogon.lateralis</i>	<i>Rynchospora.flexuosa</i>	<i>Axonopus.ciccus</i>
65	24	13

Análise 1 (as espécies são igualmente comuns no campo?)

```
#Importando os dados para o R  
qui.ex<-read.table("qui.ex.txt", h = T)
```

```
#Teste exato do qui-quadrado  
#O argumento p = c() diz ao R os valores das proporções  
chisq.test(qui.ex, p = c(1/3,1/3,1/3))
```

Análise 2 (as proporções são iguais a uma expectativa teórica de 3:2:1?)

```
chisq.test(qui.ex, p = c(3/6,2/6,1/6))
```

Os resultados dos dois testes mostram os valores de qui-quadrado (χ -squared), o valor dos graus de liberdade (df) e o valor de p (p -value).

8.6 TESTE DO QUI-QUADRADO DE INDEPENDÊNCIA

Sintaxe

`chisq.test(x)`, onde x = um vetor numérico.

Dados

Copie a tabela no editor de textos, nomeie o arquivo como `qui.ind.txt` e grave no diretório padrão.

	Jovens	Adultos
Folhas	1344	62
Ramos	1086	37

Análise

```
#Importando os dados para o R  
qui.ind<-read.table("qui.ind.txt", h = T)
```

```
#Teste exato do qui-quadrado  
#O argumento "correct=F" diz ao comando que é falsa (F=FALSE) a neces-  
sidade de correção de Yates para número pequeno de observações.  
chisq.test(qui.ind, correct=F)
```

A hipótese nula do teste é que as duas variáveis nominais são independentes. O resultado mostra os valores de qui-quadrado (χ -squared), o valor dos graus de liberdade (df) e o valor de p (p -value).

8.7 TESTE EXATO DE FISHER

Sintaxe

`fisher.test(x)`, onde x = um vetor numérico.

Dados

Copie a tabela no editor de textos, nomeie o arquivo como `fisher.txt` e grave no diretório padrão.

	Sim	Nao
Homem	15	8
Mulher	9	18

Análise

```
#Importando os dados para o R
fisher<-read.table("fisher.txt", h = T)
```

```
#Teste exato de Fisher
```

#O argumento `alternative="two.sided"` diz ao comando que realize os cálculos das probabilidades para os dois lados da distribuição.

```
fisher.test(fisher, alternative="two.sided")
```

Nossa hipótese nula é que as duas variáveis nominais são independentes. O resultado mostra o valor de p (p -value) e indica que a hipótese alternativa é que a razão das chances (*odds ratio*) verdadeira não é igual a 1. *Odds ratio* estimada é a razão das chances e mostra que a chance de homens fazerem exercício é em torno de 3,6 vezes mais do que mulheres os fazerem. *95 percent confidence interval* é o intervalo de confiança para o valor da razão das chances. Como o valor 1 (a razão das chances é igual) está fora desse intervalo, a hipótese nula é rejeitada. Note que podemos utilizar o valor de p ou o intervalo de confiança para testar a hipótese nula.

8.8 TESTE DE MCNEMAR

Sintaxe

`mcnemar.test(x)`, onde x = tabela de contingência 2×2 .

Dados

Copie a tabela no editor de textos, nomeie o arquivo como `mcnear.txt` e grave no diretório padrão.

	Correto	Incorreto
Correto	10	1
Incorreto	20	9

Análise

```
#Importando os dados para o R
#O comando row.names=1 diz ao comando que os nomes das linhas devem ser lidos
```

```
mcnemar<-read.table("mcnemar.txt", h = T, row.names=1)
```

```
#Transformando os dados em matriz
```

```
mcnemar<-as.matrix(mcnemar)
```

```
#Teste de McNemar
```

```
#O comando correct=F diz para não aplicar a correção de continuidade. Se você não inserir esse comando, a função aplicará essa correção.
```

```
mcnemar.test(mcnemar,correct=F)
```

Nossa hipótese nula é que não há diferença na resposta dos cães ao método de adestramento. O resultado mostra o valor do teste qui-quadrado de McNemar (*McNemar's chi-squared*), o valor dos graus de liberdade (*df*) e o valor de p (*p-value*).

8.9 TESTE-t

Sintaxe

`t.test(x,y)`, onde x e y = vetores numéricos.

Dados

Copie a tabela no editor de textos, nomeie o arquivo como `t.txt` e grave no diretório padrão.

A	B
175.26	172.72
177.80	157.48
167.64	170.18
160.02	172.72
172.72	175.26
177.80	170.18
175.26	154.94
170.18	149.86
157.48	157.48
160.02	154.94
193.04	175.26
149.86	167.64
157.48	157.48
157.48	157.48
190.50	154.94
157.48	177.80
182.88	160.02

Análise

```
#Importando os dados para o R
t<-read.table("t.test.txt", h = T)
```

```
#Teste-t
```

#O argumento `"var.equal=T"` diz ao comando que as variâncias dos dois grupos são iguais (TRUE), como já tínhamos visto com o teste de Bartlett.

```
t.test(t$A, t$B, var.equal=T)
```

A hipótese nula é que não há diferenças significativas nas médias de altura entre as duas classes. O resultado mostra o valor do teste-t (t), o número de graus de liberdade (df) e o valor de p (p -value). A hipótese alternativa é que a diferença real entre as médias é diferente de zero. Os valores do intervalo de confiança mostram que o zero está contido nos limites e, portanto, as duas médias são iguais e se aceita a hipótese nula. Note que podemos utilizar o valor de p ou o intervalo de confiança para testar a hipótese nula de igualdade das médias.

8.9.1 Test-t de Welch

Para realizar o teste-t de Welch, quando dois grupos possuem desvios-padrões que são substancialmente diferentes (por exemplo, um desvio padrão é três vezes maior que o outro) e os tamanhos amostrais das amostras são pequenos (menor que 10) ou diferentes, utiliza-se o mesmo comando do teste-t, somente alterando o argumento “var.equal=F”, ou seja, informando ao programa que as variâncias não são iguais.

8.10 ANOVA DE UMA VIA

Sintaxe

`aov(y ~ x, dados)`, onde x = variável X com “n” classes, y = variável quantitativa e “dados” é o nome do arquivo.

Dados

Copie a tabela no editor de textos, nomeie o arquivo como F1.txt e grave no diretório padrão.

AREAS	N
FA	64
FA	92
FA	99
PA	82
PA	129
PA	128
PP	118
PP	183
PP	196

PE	113
PE	27
PE	41

* FA = Floresta com Araucária; PA = plantação de Araucária; PP = plantação de pinos; PE = plantação de eucaliptos.

Análise

```
#Importando os dados para o R
```

```
F1<-read.table("F1.txt", h = T)
```

```
#Teste F, o comando cria um objeto chamado "anova1" e armazena o resultado do teste executado pelo comando "aov" (analysis of variance).
```

```
anova1<-aov(N ~ AREAS, F1)
```

```
#Tabela de ANOVA
```

```
summary(anova1)
```

```
#Visualizando as comparações de todas as médias (teste de contraste)
```

```
TukeyHSD(anova1)
```

```
#Estimando intervalos de confiança para as médias dos grupos
```

```
#Necessita instalar pacote emmeans
```

```
install.packages("emmeans")
```

```
library(emmeans)
```

```
emmeans(anova1, pairwise ~ AREAS)
```

Nossa hipótese nula é que não há diferenças significativas nas médias de abundância entre as quatro classes. O resultado da ANOVA mostra a tabela de resultado básica. *Df* = graus de liberdade; *Sum Sq* = soma de quadrados; *Mean Sq* = quadrados médios; *F value* = valor do teste-*F*; *Pr (>F)* = valor de *p*. O resultado mais importante do teste de contraste é o valor de *p* ajustado – *p adj*. Se esse valor for menor do que 0,05 para um contraste (comparação entre duas médias), então, rejeitamos a hipótese nula de que as médias são iguais. Os intervalos de 95% de confiança para as médias dos grupos mostram o intervalo provável de encontrar a média “real”. Lembrando que não se deve analisar a sobreposição ou não dos intervalos de confiança para decidir se duas médias são ou não iguais.

Desafio

1. Realize testes de Shapiro-Wilk e Bartlett para cada grupo da ANOVA.

8.11 ANOVA DE DUAS VIAS

Sintaxe

$\text{aov}(y \sim x*z, \text{dados})$, onde “y” = variável quantitativa, “x” = variável nominal, “z” = variável nominal, “dados” = seus dados.

Dados

Copie a tabela no editor de textos, nomeie o arquivo F2.txt e grave no diretório padrão.

DIVERSIDADE	TAMANHO	COLONIZ
Mais	Grande	0.21622
Mais	Pequena	0.266667
Mais	Grande	0.24675
Mais	Pequena	0.19231
Mais	Grande	0.19512
Mais	Pequena	0.23077
Mais	Grande	0.27451
Mais	Pequena	0.2
Mais	Grande	0.21591
Mais	Pequena	0.18868
Menos	Grande	0.20339
Menos	Pequena	0.17647
Menos	Grande	0.206897
Menos	Pequena	0.18644
Menos	Grande	0.2053
Menos	Pequena	0.18182
Menos	Grande	0.2375
Menos	Pequena	0.22727
Menos	Grande	0.19737
Menos	Pequena	0.23684

Análise

```
#Importando os dados para o R  
F2<-read.table("F2.txt", h = T)
```

#Teste-F – cria um objeto chamado “anova2” e armazena o resultado do teste executado pelo comando “aov”.

```
anova2<-aov(COLONIZ ~ DIVERSIDADE*TAMANHO, F2)
```

```
#Tabela de ANOVA  
summary(anova2)
```

```
#Intervalos de confiança e testes de contrastes  
emmeans(anova2, pairwise ~ TAMANHO*DIVERSIDADE)
```

Nossas hipóteses nulas são (1) que não haja diferenças significativas nas médias de colonização entre manchas vegetais grandes e pequenas; (2) que não haja diferenças significativas nas médias de colonização entre manchas vegetais mais e menos diversas; e (3) que não haja interação entre as variáveis nominais. Como praxe, você deve olhar primeiro para o teste da interação entre as duas variáveis nominais. Se houver interação significativa, o habitual é realizar testes individuais para cada fator separadamente com os níveis do outro fator como classes. O resultado da ANOVA mostra a tabela de resultado básica. *Df* = graus de liberdade; *Sum Sq* = soma de quadrados; *Mean Sq* = quadrados médios; *F value* = valor do teste-*F*; *Pr (>F)* = valor de *p*.

8.12 ANOVA ANINHADA

Sintaxe:

`aov(y ~ x/z, dados)`, onde *y* = variável quantitativa, *x* = variável nominal com “*n*” classes, *z* = variável nominal com “*n*” classes aninhada em *x* e “*dados*” é o nome do arquivo.

Dados

Copie a tabela no editor de textos, nomeie o arquivo como F3.txt e grave no diretório padrão.

RIPARIA	DISTANCIA	RIQUEZA
COM	1	5
COM	2	7
COM	3	15
COM	1	10
COM	2	8
COM	3	15
COM	1	6
COM	2	11
COM	3	9
SEM	1	11
SEM	2	7
SEM	3	10
SEM	1	7
SEM	2	6
SEM	3	7
SEM	1	8
SEM	2	5
SEM	3	6
SEM	1	5
SEM	2	6
SEM	3	7
SEM	1	7
SEM	2	5
SEM	3	5
SEM	1	7
SEM	2	5
SEM	3	6
COM	1	6
COM	2	6
COM	3	5
COM	1	6
COM	2	9
COM	3	4

COM	1	3
COM	2	3
COM	3	6
COM	1	9
COM	2	10
COM	3	11
COM	1	17
COM	2	6
COM	3	11
COM	1	10
COM	2	7
COM	3	9
SEM	1	6
SEM	2	9
SEM	3	4
SEM	1	4
SEM	2	5
SEM	3	4
SEM	1	6
SEM	2	2
SEM	3	4

Análise:

#Importando os dados para o R

```
F3<-read.table("F3.txt", h = T)
```

#Teste-*F* – cria um objeto chamado “anova3” e armazena o resultado do teste realizado pelo comando “aov”

```
anova3<-aov(RIQUEZA ~ RIPARIA/DISTANCIA, F3)
```

#Tabela de ANOVA

```
summary(anova3)
```

Nossas hipóteses nulas são: (1) que não haja diferenças significativas nas médias de riqueza entre zonas ripárias com e sem mata e (2) que não haja interação entre as variáveis nominais tipos de zona ripária e distância. Como praxe, você deve olhar primeiro para o teste da interação entre as duas variáveis nominais. Se houver interação significativa, o habitual é realizar testes individuais para cada fator separadamente com os níveis do outro fator como classes. O resultado da ANOVA aninhada mostra a tabela de resultado básica. *Df* = graus de liberdade; *Sum Sq* = soma de quadrados; *Mean Sq* = quadrados médios; *F value* = valor do teste-*F*; *Pr (>F)* = valor de *p*. Você pode usar o comando `emmeans` para ver os intervalos de confiança. Porém, como há uma interação aninhada, deve-se realizar ANOVAs individuais com cada fator e, depois, realizar o cálculo dos intervalos de confiança para cada fator separadamente.

8.13 CORRELAÇÃO

8.13.1 Correlação de Pearson

Sintaxe

`cor.test(x,y)`, onde “x” = variável quantitativa e “y” = variável quantitativa.

Dados

Copie a tabela no editor de textos, nomeie o arquivo como `corr.txt` e grave no diretório padrão.

PRESSAO	PESO
105	85.4
115	94.2
116	95.3
117	94.7
112	89.4
121	99.5
121	99.8
110	90.9

110	89.2
114	92.7
114	94.4
115	94.1
114	91.6
106	87.1
125	101.3
114	94.5
106	87
113	94.5
110	90.5
122	95.7

Análise

```
#Importando os dados para o R  
corr<-read.table("corr.txt", h = T)
```

```
#Teste de correlação (coeficiente de correlação de Pearson)  
cor.test(corr$PRESSAO, corr$PESO)
```

Nossa hipótese nula é que não haja associação entre as duas variáveis quantitativas. O resultado do teste mostra o valor de r na última linha (*cor*) e o valor de p (*p-value*). O resultado também indica que a hipótese alternativa é que a correlação real é diferente de zero. Mostra, também, o intervalo de confiança. Como o zero não está incluído dentro dos limites de confiança e o valor de $p < 0,05$ rejeita-se a hipótese nula. Note que se pode usar os limites de confiança ou o valor de p como critério.

8.13.2 Matriz de correlação

Dados

Copie a tabela no editor de textos, nomeie o arquivo como *corr.mat.txt* e grave no diretório padrão.

TURB	ALCA	OD	COND	TEM	PH
4.11	0.43	8.15	13.8	22	6.95
7.92	0.56	9.02	27.8	15	8.8
2.06	0.6	7.5	39.9	19	8
5.9	0.46	7.5	43.8	19	6.7
7.04	0.33	6.95	23.1	24	7.06
8.8	0.43	6.75	29.7	22	7.17
2.15	0.63	7	34.9	20	6.5
6.71	0.33	8.14	10.8	22	6.89
29.55	0.46	5.3	70.7	20	7.1
6.46	0.73	6.6	432.43	18	7.36
8.09	0.66	6	46	24	6.8
6.78	0.66	6.97	57.8	22	7.27

```
#Exportando os dados para o R
corrmat<-read.table("corr.mat.txt",h=T)
```

```
#Análise
cor(corrmat)
```

A análise mostra uma matriz com os valores dos coeficientes de correlação de Pearson (r) acima e abaixo da diagonal principal (de 1s). Para gerar os valores de p para cada correlação, deve-se realizar a análise de correlação para cada par de variável.

Desafio

1. Realize todas as correlações de Pearson par a par para identificar os valores de p .

8.13.3 Correção do valor de p

O valor de p para cada correlação par a par da matriz de correlação deve ser corrigido por causa das comparações múltiplas. Para isso, usamos um comando específico.

```
#Ajuste do valor de  $p$  da correlação entre as variáveis TURB e ALCA  
p.adjust(0.5427, method="bonferroni", n=6)
```

O primeiro valor é o valor de p sem ajuste calculado para as duas variáveis, o método pode ser Bonferroni ou outros (chame a ajuda do comando para saber as outras opções) e o “n” é o número total de variáveis na matriz de correlação.

Desafio

1. Aplique a correção de Bonferroni para todas as correlações de Pearson da matriz.

8.13.4 Correlação de Spearman

```
cor.test(corr$PRESSAO, corr$PESO, method="spearman")
```

Desafio

1. Realize a correlação de Spearman par a par para as seis variáveis da matriz de correlação.

8.14 REGRESSÃO LINEAR SIMPLES

Sintaxe

$\text{lm}(y \sim x, \text{dados})$, onde “y” = variável quantitativa dependente, “x” = variável quantitativa independente e “dados” nome do objeto no R com os dados.

Dados

O arquivo é igual ao da correlação, o nome é corr.

Análise

```
#Os dados são iguais aos da análise de correlação
```

```
#Teste
```

```
modelo<-lm(PRESSAO ~ PESO, corr)
```

```
summary(modelo)
```

Intervalo de confiança de 95% para os coeficientes
confint(modelo, level=0.95)

A hipótese nula é de que não haja associação significativa entre as duas variáveis quantitativas. O resultado da regressão mostra os coeficientes β_0 (*Intercept*) e β_1 e, no parágrafo inferior, o resultado de r^2 (*Multiple R-squared*), o r^2 ajustado (*Adjusted R-squared*), os graus de liberdade (*DF*), e o valor de p (*p-value*). Com os valores dos coeficientes, podemos montar a equação da reta. O intervalo de confiança mostra em que amplitude de valores esperamos encontrar os valores dos coeficientes da reta “real” com 95% de confiança. Quanto menor o intervalo, mais acurada será a estimativa.

Desafio

1. Monte a equação da reta de regressão. O que os valores dos coeficientes indicam na relação entre as duas variáveis?

8.15 REGRESSÃO LINEAR MÚLTIPLA

Sintaxe

$\text{lm}(y \sim x_1 + \dots + x_n, \text{dados})$, onde “y” é a variável resposta dependente, “ x_1 ” a “ x_n ” são as variáveis independentes e “dados” é o objeto do R com seus dados.

Dados

Copie a tabela no editor de textos, nomeie o arquivo como reg.multi.txt e grave no diretório padrão.

RAO	TURB	ALCA	OD	COND	TEM	PH
0.03	4.11	0.43	8.15	13.8	22	6.95
0.3	7.92	0.56	9.02	27.8	15	8.8
0.41	2.06	0.6	7.5	39.9	19	8
0.24	5.9	0.46	7.5	43.8	19	6.7
0.13	7.04	0.33	6.95	23.1	24	7.06
0.22	8.8	0.43	6.75	29.7	22	7.17
0.46	2.15	0.63	7	34.9	20	6.5
0.074	6.71	0.33	8.14	10.8	22	6.89
0.28	29.55	0.46	5.3	70.7	20	7.1

0.35	6.46	0.73	6.6	432.43	18	7.36
0.26	8.09	0.66	6	46	24	6.8
0.34	6.78	0.66	6.97	57.8	22	7.27

Análise

```
#Importando os dados para o R
reg.multi<-read.table("reg.multi.txt",h=T)
```

```
# Teste de regressão múltipla
reg.multi.res<-lm(RAO ~ TURB+ALCA+OD+COND+TEM+PH, reg.multi)
summary(reg.multi.res)
```

```
# Intervalos de confiança
confint(reg.multi.res, level = 0.95)
```

Nossa hipótese nula é de que não há associação entre a variável dependente e nenhuma das variáveis independentes. O resultado da análise é uma tabela com as estimativas dos efeitos de cada variável e do intercepto, o erro-padrão (*Std.Error*), valor de t (t value) e o valor de p ($Pr[>|t|]$). No parágrafo abaixo da tabela, são mostrados os graus de liberdade (*degrees of freedom*), o R^2 (*Multiple R-squared*), o valor de R^2 ajustado (*Adjusted R-squared*), o teste- F (*F-statistic*) e o valor de p (p -value). Os intervalos de confiança mostrados são os do intercepto e os dos coeficientes individuais das variáveis explicativas. A interpretação é similar à realizada para o modelo de regressão linear simples.

8.16 REGRESSÃO LINEAR MÚLTIPLA COM SELEÇÃO DE VARIÁVEIS

8.16.1 Seleção para frente ajustada

```
#Instalar pacote "adespatial"
install.packages("adespatial")
library(adespatial)
```

```
#Separar a variável dependente das independentes no arquivo reg.multi
RAO<-reg.multi[,1]
X<-reg.multi[,2:7]
```

```
#Realizar a análise
forward.sel(RAO,X)
```

O comando realizará os testes e mostrará o resultado com a(s) variável(is) escolhidas, o valor de R^2 , o valor de R^2 acumulado ($R2cum$), o valor de R^2 acumulado ajustado ($AdjR2Cum$), o valor do teste- F (F) e o valor de p ($pval$).

8.16.2 Seleção AIC

```
#Criando um modelo nulo (sem nenhuma variável, só o intercepto)
nulo<-lm(RAO ~ 1, reg.multi)
```

```
#Criando um modelo completo (todas as variáveis)
completo<-lm(RAO ~ TURB+ALCA+OD+COND+TEM+PH, reg.multi)
```

```
#Realizando a análise de seleção de modelo
step(nulo, scope=list(upper=completo, direction="both"))
```

O resultado mostra o valor de AIC para o modelo nulo e para cada uma das variáveis em uma tabela. O melhor modelo reduzido é o que apresenta o menor valor de AIC e, na tabela, são mostrados os valores de graus de liberdade (Df), da soma de quadrados ($Sum\ of\ Sq$), dos resíduos das somas de quadrados (RSS) e do AIC . Depois, vai mostrando os modelos a partir da seleção das variáveis.

8.17 ANCOVA

Sintaxe

$aov(y \sim z*x, dados)$, onde “y” = variável quantitativa, “z” = variável quantitativa covariável, “x” = variável nominal com “n” classes e “dados” é o nome do arquivo.

Dados

Copie a tabela no editor de textos, nomeie o arquivo como ancova.txt e grave no diretório padrão.

Duracao	Ano	Colheita
36	2006	67.26
50	2006	60.34
64	2006	53.69
78	2006	52.54
92	2006	53.8
106	2006	60.69
0	2006	67.07
106	2006	33.01
92	2006	39.03
78	2006	49.47
64	2006	53.48
50	2006	58.55
36	2006	63.64
0	2006	70.81
64	2006	69.7
36	2006	65.58
106	2006	55.47
50	2006	67
92	2006	53.5
78	2006	64.69
0	2006	71.38
50	2006	67.79
78	2006	73.32
92	2006	61.4
36	2006	62.48
106	2006	44.74
64	2006	56.02
0	2006	65.12
133	2006	40.83
133	2006	40.03

133	2006	34.05
133	2006	50.94
36	2007	47.88
50	2007	49.21
64	2007	43.89
78	2007	49.21
92	2007	33.25
106	2007	29.26
78	2007	49.21
50	2007	54.53
106	2007	34.58
36	2007	73.15
64	2007	39.9
92	2007	34.58
92	2007	45.22
106	2007	49.21
36	2007	57.19
78	2007	53.2
50	2007	50.54
64	2007	65.17
78	2007	39.9
92	2007	43.89
64	2007	65.17
106	2007	31.92
36	2007	49.21
50	2007	65.17
0	2007	65.17
0	2007	62.51
0	2007	55.86
0	2007	65.17
133	2007	35.91
133	2007	37.24
133	2007	45.22
133	2007	34.58

Análise

```
#Importando os dados para o R
```

```
ancova<-read.table("ancova.txt", h = T)
```

```
#Com comando "aov"
```

```
#Analisando se há interação entre as variáveis explicativas
```

```
ancova.int<-(aov(Colheita ~ Duracao*Ano, ancova))
```

```
summary(ancova.int)
```

```
#Analisando os efeitos individuais das variáveis explicativas
```

```
ancova.ind<-(aov(Colheita ~ Duracao+Ano, ancova))
```

```
summary(ancova.ind)
```

```
#Com comando "lm"
```

```
#Analisando se há interação entre as variáveis explicativas
```

```
ancova.int.lm<-(lm(Colheita ~ Duracao*Ano, ancova))
```

```
summary(ancova.int.lm)
```

```
# Analisando os efeitos individuais das variáveis explicativas
```

```
ancova.ind.lm<-(lm(Colheita ~ Duracao+Ano, ancova))
```

```
summary(ancova.ind.lm)
```

```
# Intervalo de confiança para as variáveis Duracao e Ano
```

```
confint(ancova.ind, 'Duracao',level=0.95)
```

```
confint(ancova.ind,'Ano',level=0.95)
```

Como uma ANCOVA é uma análise híbrida com elementos de ANOVA e regressão, ela pode ser realizada no R tanto com o comando "aov", quanto com o comando "lm". Os resultados são ligeiramente diferentes, pois o resultado do comando "aov" mostra uma tabela de ANOVA, enquanto o resultado do comando "lm" mostra os coeficientes de regressão, o valor de r^2 e outros resultados. Porém, os valores de p são os mesmos. Os resultados do comando "lm" mostram o sentido da variação na variável resposta por influência da covariável, mas isso não é mostrado nos resultados do comando "aov". Note que o resultado de uma análise ("summary") pode ser chamado diretamente, sem precisar gravar em um objeto novo. Os intervalos de confiança são interpretados da mesma forma que nos resultados da ANOVA para a variável Ano, e da regressão linear simples para a variável Duracao.

8.18 REGRESSÃO LOGÍSTICA

Sintaxe

`glm(y ~ x, family=binomial(link="logit"))`, onde “y” = variável nominal, “x” = variável quantitativa, “family” é a família da distribuição de frequência e “logit” é o chamado para a análise de regressão logística.

Dados

Copie a tabela no editor de textos, nomeie o arquivo de log.txt e grave no diretório padrão.

Area	Presenca
3.798507	1
2.373975	0
11.99711	0
6.864291	0
4.901955	1
0.830087	1
0.363559	1
6.260152	1
13.69452	0
0.606617	1
2.759371	0
1.069333	0
1.259256	1
2.395813	1
1.286634	0
1.053435	1

Análise

```
#Importando os dados para o R  
log<-read.table("log.txt", h = T)
```

```
#Regressão logística tradicional com valor de p padrão do sistema  
reg.log<-(glm(Presenca ~ Area, family=binomial(link="logit"),log))
```

```
#Cálculo do valor de  $p$  com razão de verossimilhança  
modelo1<- glm(Presenca ~ Area, family=binomial(link="logit"),log)  
modelo2<-glm(Presenca ~ 1, family=binomial(link="logit"),log)
```

```
#Instalar pacote "lmtree" e fazer o teste  
install.packages("lmtree")  
library(lmtree)  
lmtree(modelo1, modelo2)
```

```
# Intervalo de confiança  
confint(reg.log, level=0.95)
```

Nossa hipótese nula é de que não há relação entre a variável nominal e a variável quantitativa. O resultado mostra uma tabela com os coeficientes dos efeitos do intercepto (intercept) e da variável quantitativa, o erro-padrão (Std. Error), o valor de z (z value) e o valor de P ($\Pr[>|z|]$). Os valores de "deviance" são equivalentes aos valores de soma de quadrados da regressão linear. No caso do intervalo de confiança, se ele contém o número 1, então aceita-se a hipótese nula, o que é o caso para o nosso exemplo.

REFERÊNCIAS

- AKAIKE, H. Information theory and the extension of the maximum likelihood principle. *In*: PARZEN, E.; TANABE, K.; KITAGAWA, G. (ed.). **Selected papers of Hirotugu Akaike**. Springer Series in Statistics (Perspectives in Statistics). Springer: New York, 1998. p. 199-213.
- BALDISSERA, R.; GANADE, G.; BRESCOVIT, A. D.; HARTZ, S. M. Landscape mosaic of *Araucaria* forest and forest monocultures influencing understorey spider assemblages in southern Brazil. **Austral Ecology**, v. 33, n. 1, p. 45-54, 2008a.
- BALDISSERA, R.; FRITZ, L.; RAUBER, R.; MÜLLER, S. C. Comparison between grassland communities with and without disturbances. **Neotropical Biology and Conservation**, v. 5, n. 1, p. 3-9, 2010.
- BALDISSERA, R.; BACH, E.; LIMA, R. P. DE; MENEGASSI, A.; PIOVESAN, A. R.; FONSECA, G. C. DA. Distribuição de aranhas de teia de sub-bosque ao longo de uma área de contato entre floresta de *Araucaria* e plantação de *Pinus* no sul do Brasil. **Neotropical Biology and Conservation**, v. 3, n. 1, p. 3-8, 2008b.
- BALDISSERA, R. **Estrutura e composição de assembleias de aranhas em manchas de vegetação na porção austral da Mata Atlântica**. 2012. 147 p. Tese (Doutorado em Ciências – ênfase em Agroecologia) – Instituto de Biociências, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, 2012.
- BALDISSERA, R.; RODRIGUES, RODRIGUES, E. N. L.; HARTZ, S. M. Assessment of the probability of colonization of local spider communities in an experimental landscape. **The Journal of Arachnology**, v. 41, n. 2, p. 160-167, 2013.
- BLANCHET, F. G.; LEGENDRE, P.; BORCARD, D. Forward selection of explanatory variables. **Ecology**, v. 89, n. 9, p. 2623-2632, 2008.
- BONFERRONI, C. E. Il calcolo delle assicurazioni su gruppi di teste. **Studi in Onore del Professore Salvatore Ortu Carboni**. Roma, Itália, 1935. p. 13-60.
- BONNINI, S.; CORAIN, L.; MAROZZI, M.; SALMASO, L. **Nonparametric hypothesis testing: rank and permutation methods with applications in R**. Chichester, Reino Unido: John Wiley & Sons, 2014. 254 p.
- BUSSAB, W. DE O.; MORETTIN, P. A. **Estatística básica**. São Paulo: Saraiva, 2010.

CAIN, J. W. Mathematical models in science. In: BELL, E. (ed.). **Molecular Life Sciences**. Springer: New York, 2014.

COHEN, J. **Statistical power analysis for the behavioral sciences**. New Jersey: Lawrence Erlbaum Associates Publishers, 1988. 567 p.

CONNELL, J. H.; SLATYER, R. O. Mechanisms of succession in natural communities and their role in community stability and organization. **The American Naturalist**, v. 111, n. 982, p. 1119-1144, 1977.

CORÁ, D. H.; OLIVEIRA, A. D. de; GALIANO, D. **Variação na abundância de sturnira liliium (chiroptera: phyllostomidae) em relação à disponibilidade de diásporos zoocóricos em um fragmento urbano de Mata Atlântica**. Monografia (conclusão do curso de ciências biológicas - bacharelado) - Universidade Comunitária da Região de Chapecó, 2018.

CRAWLEY, M. J. **The R book**. West Sussex, England: John Wiley & Sons, 2007. 951 p.

DIAMOND, J. M. Ecology: laboratory, field and natural experiments. **Nature**, v. 304, p. 586-587, 1983.

DIENES, Z. Bayesian versus orthodox statistics: which side are you on? **Perspectives on Psychological Science**, v. 6, n. 3, p. 274-290, 2011.

DU PREL, J.-B.; HOMMEL, G.; RÖHRIG, B.; BLETTNER, M. Confidence interval or p-value? **Deutsches Ärzteblatt International**, v. 106, n. 19, p. 335-339, 2009.

EDGELL, S. E.; NOON, S. M. Effect of violation of normality on the *t* test of the correlation coefficient. **Psychological Bulletin**, v. 95, n. 3, p. 576-583, 1984.

ELLIS, P. D. **The essential guide to effect sizes: statistical power, meta-analysis, and the interpretation of research results**. Cambridge: Cambridge University Press, 2010. 193 p.

ESTUDO. **Michaelis Moderno Dicionário da Língua Portuguesa**. São Paulo: Melhoramentos. Disponível em: <https://michaelis.uol.com.br/moderno-portugues/>.

GALETI, G. **Diversidade taxonômica e funcional de aranhas edáficas em matas ciliares do bioma Mata Atlântica**. 2020. 82p. Dissertação (Mestrado em Ciências Ambientais) – Universidade Comunitária da Região de Chapecó (Unochapecó), Chapecó-SC, 2020.

GLASS, G. V.; PECKHAM, P. D.; SANDERS, J. R. Consequences of failure to meet assumptions underlying fixed effects analyses of variance and covariance. **Review of Educational Research**, v. 42, n. 3, p. 237-288, 1972.

GREENLAND, S.; SENN, S. J. ROTHMAN, K. J.; CARLIN, J. B.; POOLE, C.; GOODMAN, S. N.; ALTMAN, D. G. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. **European Journal of Epidemiology**, v. 31, p. 337-350, 2016.

HAND, D. J. **Statistics: a very short introduction**. Oxford: Oxford University Press, 2008.

HARWELL, M. R.; RUBINSTEIN, E. N.; HAYES, W. S.; OLDS, C. C. Summarizing Monte Carlo results in methodological research: the one- and two-factor fixed effects ANOVA cases. **Journal of Educational Statistics**, v. 17, n. 4, p. 315-339, 1992.

HURLBERT, S. H. Pseudoreplication and the design of ecological field experiments. **Ecological Monographs**, v. 54, n. 2, p. 187-211, 1984.

JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. **An introduction to statistical learning**: with applications in R. (Springer Texts in Statistics Book Series). Springer: New York, 2021. 616 p.

KROKHMAL, V. Introductory probability and the central limit theorem. 11 p. Disponível em: <http://www.math.uchicago.edu/~may/VIGRE/VIGRE2011/REUPapers/Krokhmal.pdf>. 2011. Acesso em: 16 abr. 2022.

LANZANTE, J. R. A cautionary note on the use of error bars. **Journal of Climate**, v. 18, n. 17, p. 3699-3703, 2005.

LEGENDRE, P.; LEGENDRE, L. **Numerical Ecology**. 3. ed. Holanda, Amsterdã: Elsevier, 2012.

KEPPELER JUNIOR, R.; BALDISSERA, R.; FRANCO, G. M. DE S. **Resposta funcional de comunidades de peixes a variáveis locais em riachos subtropicais no sul do Brasil**. 206. 19 p. Monografia (Curso de ciências biológicas) – Universidade Comunitária da Região de Chapecó (Unochapecó), Chapecó-SC, 2016.

KIM, J. How to choose the level of significance: a pedagogical note. **Munich Personal RePEc Archive**, n. 66373, 2015. Disponível em: <https://mpra.ub.uni-muenchen.de/66373/>.

LEEDY, P. D.; ORMROD, J. E. **Practical research**: planning and design. Essex, United Kingdom: Pearson Education, 2021.

LIX, L. M.; KESELMAN, J. C.; KESELMAN, H. J. Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance F test. **Review of Educational Research**, v. 66, n. 4, p. 579-619, 1996.

MCCULLAGH, P.; NELDER, J.A. **Generalized linear models**. Chapman and Hall: New York, London, 1989. 532 p.

MAKAR, K; RUBIN, A. Learning about statistical inference. In: Ben-Zvi D., Makar K., Garfield J. (ed.). **International Handbook of Research in Statistics Education**. Springer International Handbooks of Education. Alemanha: Springer, Cham., 2017. p. 261-294.

MANLY, B. F. J. **Randomization, bootstrap and Monte Carlo methods in biology**. Boca Raton, Flórida: Chapman & Hall/CRC, 2006. 468 p.

MCDONALD, J. H. **Handbook of biological statistics**. Baltimore, Maryland: Sparky House Publishing, 2014. 305 p.

MELLO, M. P.; PETERNELLI, L. A. **Conhecendo o R: uma visão mais que estatística**. Viçosa: UFV, 2013. 222 p.

MENDENHALL, III, W.; BEAVER, R. J.; BEAVER, B. M. **Introduction to probability and statistics: metric version**. USA: Cengage, 2018.

NARUM, S. R. Beyond Bonferroni: less conservative analysis for conservation genetics. **Conservation Genetics**, v. 7, p. 783-787, 2006.

NUZZO, R. Scientific method: Statistical errors. **Nature**, v. 506, p. 150-152, 2014.

ODERO, D. C.; MESBAH, A. O.; MILLER, S. D.; KNISS, A. R. Lanceleaf (*Salvia reflexa*) interference in sugarbeet. **Weed Technology**, v. 24, n. 4, p. 557-661, 2010.

OEHLERT, G. W. **A first course in design and analysis of experiments**. University of Minnesota, 2010. 679 p.

OESTREICHER, C. A history of chaos theory. **Dialogues in Clinical Neuroscience**, v. 9, n. 3, p. 279-289, 2007.

PAULINO, C. D.; TURKMAN, M. A. A.; MURTEIRA, B.; SILVA, G. L. **Estatística Bayesiana**. Lisboa, Portugal: Fundação Calouste Gulbekian, 2018. 601 p.

PENN STATE UNIVERSITY. **Stat 462 Applied Regression Analysis online course**. Disponível em: online.stat.psu.edu/stat462/. 2018. Acesso em: 22 out. 2020.

PERKINS, J.; WANG, D. A comparison of Bayesian and Frequentist statistics as applied in a simple repeated measures example. **Journal of Modern Applied Statistical Methods**, v. 3, n. 1, p. 227-233, 2004.

PESQUISA. **Michaelis Moderno Dicionário da Língua Portuguesa**. São Paulo: Melhoramentos. Disponível em: <https://michaelis.uol.com.br/moderno-portugues/>.

QUINN, G. P.; KEOUGH, M. J. **Experimental design and data analysis for biologists**. Cambridge: Cambridge University Press, 2002. 553 p.

R Core Team. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. 2020. Disponível em: <http://www.r-project.org/index.html> <https://www.R-project.org/>. Acesso em: 16 abr. 2022.

RASCH, D.; GUIARD, V. The robustness of parametric statistical methods. **Psychology Science**, v. 46, n. 2, p. 175-208. 2004.

REBESCHINI, C.; GALETI, G.; DE FARIAS, P. M.; BALDISSERA, R. Edaphic beetle (Coleoptera) community response to the presence of gallery forest in a Brazilian subtropical grassland landscape. **Biologia**, v. 9, p. 2503-2510, 2021.

REID, H. M. **Introduction to statistics: fundamental concepts and procedures of data analysis**. Los Angeles, USA: Sabe Publications, 2013.

ROSS, D. Game Theory. **The Stanford Encyclopedia of Philosophy** (Fall 2021 Edition). Edward N. Zalta (ed.). Disponível em: <https://plato.stanford.edu/archives/fall2021/entries/game-theory/>. Acesso em: 16 abr. 2022.

SCHEINER, S. M.; GUREVITCH, J. **Design and analysis of ecological experiments**. Nova Iorque: Oxford University Press, 2001. 415 p.

SILVA, H. A. da. **Manual básico da linguagem R: introdução à análise de dados com a linguagem R e o RStudio para área da saúde**. [Recurso eletrônico] São João del Rei: Editora do Autor, 2018. 178 p.

SOKAL, R. R.; ROHLF, F. J. **Introduction to biostatistics**. New York: W. H. Freeman, 2009. 374 p.

SUE, V. M.; RITTER, L. A. **Conducting online surveys**. Thousand Oaks, California, USA: Sage Publications, 2007.

TOFT, S. Nutritional aspects of spider feeding. *In*: Nentwig W. (ed.). **Spider ecophysiology**. Springer: Berlim, 2013. p. 373-384.

WASSERSTEIN, R. L.; LAZAR, N. A. The ASA's statement on p-values: context, process, and purpose. **The American Statistician**, v. 70, n. 2, p. 129-133, 2016.

ZHANG, J.; HANIK, B. W.; CHANEY, B. H. Confidence intervals: evaluating and facilitating their use in health education research. **The Health Educator**, v. 40, n. 1, p. 29-39, 2008.

Aleatorização: procedimento de escolha de objetos de um universo amostral no qual todos os objetos possuem a mesma probabilidade de serem escolhidos.

Amostra: subconjunto de um universo amostral.

Atribuição aleatória: procedimento de atribuição aleatória de objetos para grupos experimentais.

Controle estatístico: procedimento para retirar o efeito de variáveis de confusão sobre a variável resposta feita na etapa de análise de dados.

Curva normal: representação de uma distribuição de probabilidade contínua que é simétrica nos dois lados da média.

Dependência espacial: quando uma medida de um descritor em um objeto está associada com a medida do descritor em outro objeto próximo espacialmente do primeiro.

Desenho balanceado: número igual de unidades experimentais em todos os grupos de um experimento.

Desenho experimental: conjunto de procedimentos lógicos para a coleta de dados biológicos e ambientais.

Distribuição de frequências: distribuição da frequência de ocorrência dos valores de uma variável quantitativa ao longo da amplitude de variação dos dados.

Escala de Likert: respostas ordenadas em uma escala subjetiva de concordância com uma pergunta.

Experimento: trabalho científico que tem por objetivo estudar um fenômeno biológico ou ambiental.

Falso negativo: aceitação da hipótese nula quando ela é falsa. Erro Tipo II.

Falso positivo: rejeição da hipótese nula quando ela é verdadeira. Erro Tipo I.

Frequência relativa: número de ocorrências de um evento qualquer dividido pelo número de ocorrências de todos os eventos possíveis no conjunto.

Graus de liberdade: número de categorias independentes (tamanho amostral) menos o número de parâmetros estatísticos a serem avaliados em um teste.

Hipótese de estudo: expectativa teórica sobre o funcionamento do sistema de estudo.

Hipótese estatística alternativa: hipótese de que há relação ou diferença estatística entre duas ou mais variáveis.

Hipótese estatística nula: hipótese de que não há relação ou diferença estatística entre duas ou mais variáveis.

Independência de observações: a medida de um descritor em um objeto não interfere no resultado da medida do descritor em outro objeto.

Multicolinearidade: forte correlação entre variáveis quantitativas independentes.

Nível de significância: medida da força da evidência que os resultados devem apresentar antes de rejeitar a hipótese nula. Probabilidade de rejeitar a hipótese nula quando ela é verdadeira.

Objeto de estudo: unidade de interesse do pesquisador onde serão medidos os descritores durante o experimento. Unidade experimental. Unidade amostral.

Outlier: ponto fora da curva. Observação com um valor extremo que se destaca dos outros dentro de um grupo (ANOVA) ou em um gráfico de associação entre variáveis (correlação/regressão).

Pareamento: procedimento em que o pesquisador mede o descritor em um mesmo objeto separado por espaço ou tempo com o objetivo de diminuir a variação natural do descritor.

Permutação: tipo de teste de significância estatística no qual a distribuição do teste sob a hipótese nula é obtido pelo cálculo de todos os valores possíveis do teste sob todas as combinações possíveis das observações.

População estatística: conjunto de todos os objetos passíveis de serem amostrados. Universo amostral.

Pressuposto estatístico: características que os dados devem apresentar para usar determinados testes estatísticos.

Previsão de estudo: expectativa sobre a resposta operacional ou quantitativa das variáveis do estudo.

Pseudorrepetição: quando unidades experimentais estão sob efeito de um mesmo fator, o que faz com que haja uma quebra do pressuposto de independência.

Tamanho amostral: número total de unidades experimentais/amostrais ou objetos do estudo.

Tamanho do efeito: medida quantitativa do tamanho da diferença nas médias de um descritor entre dois ou mais grupos ou da correlação e/ou dependência entre duas ou mais variáveis.

Teste de hipótese: teste que visa identificar se a probabilidade da resposta de um descritor é devida ao acaso ou a um efeito ligado a uma ou mais variáveis.

Teste paramétrico: teste estatístico que possui os pressupostos básicos de independência, normalidade e homocedasticidade.

Unidade experimental: objeto de estudo. Unidade amostral.

Universo amostral: população estatística.

Valor de p : probabilidade de conseguir o resultado observado, ou algo mais extremo, se a hipótese nula é verdadeira.

Variável: um descritor biológico ou ambiental.

Variável de confusão: descritor que não é medido e pode influenciar a resposta da variável de interesse do estudo.

Variável dependente: descritor que muda em resposta à variação de outro descritor.

Variável independente: descritor que varia independente de outro descritor.

Variável nominal: descritor nominal que assume poucos valores registrados como nomes.

Variável ordinal: descritor seminumérico ordenado.

Variável quantitativa: descritor numérico ordenado.

Variável quantitativa contínua: descritor numérico ordenado que pode assumir valores inteiros e fracionários.

Variável quantitativa discreta: descritor numérico ordenado que ocorre somente em valores inteiros.

Teorema do limite central: proposição que afirma que quando o tamanho de uma amostra aumenta, a distribuição amostral de sua média aproxima-se, cada vez mais, de uma distribuição normal.

SOBRE O AUTOR

Licenciado em Ciências Biológicas pela Universidade do Vale do Rio dos Sinos (UNISINOS). Mestre e Doutor em Ecologia, pela Universidade Federal do Rio Grande do Sul (UFRGS), onde também realizou pós-doutoramento. Lecionou durante 9 anos na graduação de cursos de Ciências Biológicas nas disciplinas de Ecologia, Metodologia Científica, Comportamento Animal e Análise de Dados. Também lecionou durante 5 anos na pós-graduação de curso de Ciências Ambientais nas disciplinas de Ecologia, Metodologia Científica e Análise de Dados. Durante esse tempo, coordenou e participou de vários projetos de pesquisa na área ambiental, tendo orientado trabalhos de conclusão de curso, especialização e dissertações de mestrado. Publicou 25 artigos científicos nas áreas da Ecologia e das Ciências Ambientais. Atualmente, leciona Biologia na Escola de Educação Básica Araújo Figueiredo e no Colégio Santa Clara, município de Urubici, Santa Catarina, onde reside com sua companheira, seus cães e seus gatos. É cofundador do Espaço ParaLlelo, coletivo que se propõe a resgatar a alma humanitária pela promoção de cursos e vivências nas áreas da arte, filosofia, ciência e terapias integrativas e complementares.

Reitor	Marcelo Recktenvald
Vice-Reitor	Gismael Francisco Perin
Chefe do Gabinete do Reitor	Rafael Santin Scheffer
Pró-Reitor de Administração e Infraestrutura	Charles Albino Schultz
Pró-Reitor de Assuntos Estudantis	Nedilson Lauro Brugnera
Pró-Reitor de Gestão de Pessoas	Claunir Pavan
Pró-Reitora de Extensão e Cultura	Patricia Romagnoli
Pró-Reitor de Graduação	Jeferson Saccol Ferreira
Pró-Reitor de Pesquisa e Pós-Graduação	Clevison Luiz Giacobbo
Pró-Reitor de Planejamento	Everton Miguel da Silva Loreto
Secretário Especial de Laboratórios	Edson da Silva
Secretário Especial de Obras	Fábio Correa Gasparetto
Secretário Especial de Tecnologia e Informação	Ronaldo Antonio Breda
Procurador-Chefe	Rosano Augusto Kammers
Diretor do Campus Cerro Largo	Bruno Munchen Wenzel
Diretor do Campus Chapecó	Roberto Mauro Dall'Agnol
Diretor do Campus Erechim	Luís Fernando Santos Corrêa da Silva
Diretor do Campus Laranjeiras do Sul	Martinho Machado Júnior
Diretor do Campus Passo Fundo	Julio César Stobbe
Diretor do Campus Realeza	Marcos Antônio Beal
Diretor da Editora UFFS	Antonio Marcos Myskiw
Chefe do Departamento de Publicações Editoriais	Marlei Maria Diedrich
Assistente de Administração	Fabiane Pedroso da Silva Sulsbach



Conselho Editorial

Alcione Aparecida de Almeida Alves	Aline Raquel Müller Tones
Antonio Marcos Myskiw (Presidente)	Sergio Roberto Massagli
Everton Artuso	Carlos Alberto Cecatto
Helen Treichel	Cristiane Funghetto Fuzinatto
Janete Stoffel	Stomara Aparecida Marques
Joice Moreira Schmalfluss	Gelson Aguiar da Silva Moser
Jorge Roberto Marcante Carlotto	Athany Gutierrez
Liziara da Costa Cabrera	Iara Denise Endruweit Battisti
Marcela Alvares Maciel	Alexandre Mauricio Matiello
Maude Regina de Borba	Claudia Simone Madruga Lima
Melissa Laus Mattos	Luiz Felipe Leão Maia Brandão
Nilce Scheffer	Geraldo Ceni Coelho
Tassiana Potrich	Andréia Machado Cardoso
Tatiana Champion	Fabiana Elias
Valdir Prigol (Vice-presidente)	Angela Derlise Stübe



Associação Brasileira
das Editoras Universitária:

Revisão dos textos	Autor
Preparação e revisão final	Marlei Maria Diedrich
Projeto gráfico e capa	Paolo Malorgio Studio
Diagramação	Paolo Malorgio Studio
Divulgação	Diretoria de Comunicação Social
Formatos	e-Pub, Mobi, PDF e Impresso

Livro produzido com apoio da Fundação Araucária, através do Programa de apoio a publicações científicas – fortalecimento de editoras (Chamada pública 23/2018).

Coordenador do projeto Antonio Marcos Myskiw

B177i Uma iniciação aos testes estatísticos para dados biológicos / Ronei Baldissera. – Chapecó : Ed. UFFS, 2022.

ISBN: 978-65-86545-89-0 (EPUB)
978-65-86545-92-0 (PDF)
978-65-86545-90-6 (MOBI)

1. Ciências da vida. 2. Pesquisa. 3. Estatística. I. Baldissera, Ronei. II. Título.

CDD: 574.5248072

Ficha catalográfica elaborada pela
Divisão de Bibliotecas – UFFS
Vanusa Maciel
CRB - 14/1478

