



Fundamental Data Analysis Tools and Concepts for Bioacoustical Research

9

Chandra Salgado Kent, Tiago A. Marques, and Danielle Harris

9.1 Introduction

Bioacoustics has emerged as a prominent, non-invasive, and innovative approach to obtaining scientific knowledge about animal behavior and ecology. As a consequence, bioacousticians play an important role in today's societies, often informing decision-makers in governments, industries, and communities. As an example, bioacousticians are often asked whether a species, a population, a community, or individual animals will sustain impacts from noise—or any other impact, of course, but noise is particularly relevant to the running theme of the book—generated from particular human activities.

C. Salgado Kent (✉)

Centre for Marine Science and Technology, Curtin University, Perth, WA, Australia

Oceans Blueprint, Perth, WA, Australia

Centre for Marine Ecosystems Research, School of Science, Edith Cowan University, Perth, WA, Australia
e-mail: c.salgadokent@ecu.edu.au

T. A. Marques

Centre for Research into Ecological and Environmental Modelling, University of St Andrews, St Andrews, Fife, UK

Departamento de Biologia Animal, Centro de Estatística e Aplicações, Faculdade de Ciências da Universidade de Lisboa, Lisbon, Portugal

D. Harris

Centre for Research into Ecological and Environmental Modelling, University of St Andrews, St Andrews, Fife, UK

Sometimes, government regulators require “yes” or “no” answers to these questions. A knowledgeable bioacoustician, any scientist in fact, will know that usually it is difficult to provide simple ‘yes’ or ‘no’ answers. This is because the magnitude of impact that is biologically significant is usually not known. For instance, imagine the question relates to whether loud construction works will result in a decline of a local population of animals. The observed impact is that animals reduce the time spent feeding. Therefore, the required reduction in time feeding that will lead to a population decline must be known to be able to provide a “yes” or “no” answer. Consequently, the bioacoustician's question is not whether there is simply a statistically significant effect, which by itself may be meaningless and even misleading (e.g., Wasserstein et al. 2019), but whether the magnitude of the effect is biologically important. That is a much more difficult question to answer, and hence why it is often ignored albeit inadvertently. By ensuring that research questions have biological relevance, bioacousticians can design studies that can draw meaningful conclusions about animals and their populations.

Once the biologically relevant question has been identified, the bioacoustician can determine what study design is required and whether it is possible to carry it out. All too commonly, constraints occur in available budgets and time allocated to undertake the research. This often results in sub-optimal study designs and sample sizes (e.g., reduced numbers of surveys, available

acoustic instruments, and/or surveyed animals). The reality is that for a bioacoustician to be able to confidently answer research questions, budgets must allow for robust experimental designs and sufficient time to collect sample sizes representative of the study population. Even when budgets and time allow for carefully designed experiments, however, environmental conditions and study animals often cannot be controlled, particularly when studied in their natural environment. Moreover, many studies occur opportunistically and are not the result of an experimental design developed specifically for the study aims. They are observational in nature and can take advantage of large, long-term existing datasets or unexpected opportunities to collect field data. In fact, data collected opportunistically are prevalent in bioacoustical studies, as many researchers take recording systems into the field during other work to use when time permits.

The challenges described above, from ensuring that the research questions have biological relevance, to evaluating the achievability of a study and reliability of its outcomes, are only a few of many challenges faced by bioacousticians. To overcome these challenges, bioacousticians must have solid foundational knowledge about the quantitative aspects of their research: from how to formulate quantitative research questions, to designing robust studies and undertaking suitable analyses. Only by having these skills can reliable conclusions and scientific claims be made.

Today, not only are there a wide range of analytical tools available to select from, but this ever-increasing number has been evolving quickly over recent decades due to the dramatic improvement in computer capacity. Moreover, ongoing research in statistics continually updates our knowledge on the suitability of commonly used methods (Wilcox 2010). In some instances, methods previously used over a wide range of applications may now only be acceptably applied to certain scenarios, with new methods superseding old ones. Having said this, while a new method may be considered the ‘Rolls Royce’ of analyses, sometimes an older, simpler

approach may still do the job well. Consequently, not only is it important for researchers to have a solid foundation in long-established analytical approaches, but they must keep up to date with new developments. In general, a researcher should understand the fundamentals involving randomness, variability, and statistical modeling discussed in this chapter, and be able to adapt them to their specific context—this understanding is arguably more valuable than a book of recipes that tells a researcher which method to use and when.

A consequence of the many advancements over recent years and the large range of analytical approaches available today is that selecting the right tool can be an overwhelming task. In fact, the right tool might not exist for a specific setting. In such cases, collaboration with an applied statistician may be fundamental. This chapter aims to give general guidance on considerations that bioacousticians should make when tasked with undertaking research resulting in what are often complex and messy bioacoustical datasets. The information presented in this chapter is by no means meant to provide a menu of analytical tools, their mathematical basis, or conditions of use. There are a large number of widely available textbooks that do just that, and many are referenced here. Bioacousticians should consult the relevant textbooks for in-depth knowledge of approaches, their applications, limitations, and assumptions about the characteristics of the data that must be met. Rather, the focus of this chapter is to provide practical guidance on: (1) the development of meaningful research questions, (2) data exploration and experimental design considerations (also see Chap. 3), and (3) common analytical approaches used today. The approach taken in this chapter is to define basic terms and concepts as they appear in the text, so that readers new to the subject can also understand the more complex concepts discussed, regardless of their prior statistical knowledge.

Note that this chapter has been written from the perspective of a biologist faced with the challenges common to bioacoustical research. If, from this chapter, the reader gains an appreciation of limitations in their data, considerations they

should make when selecting analytical approaches, and the biological relevance of their analytical outputs, then this chapter has achieved its purpose. Entire books could be written about how a bioacoustician, in fact, any ecologist, might become more quantitative. A good example of such a book is suitably named *How to be a quantitative ecologist* (Matthiopoulos 2010), which we wholeheartedly recommend as good reading after this chapter.

9.2 Developing a Clear Research Question

At the concept stage of any study, the purpose and specific research aim must be clearly defined. The research aim should be novel (i.e., not already answered in previous research). Once the general aim has been defined, the specific analytical research question can be developed. While developing the question may seem to be a simple, self-evident task, it requires careful consideration. The structure of the question drives the experimental design and selection of analytical tools, thus its accurate development is essential. To frame a question in clear, concise analytical terms, it is useful to identify the type of study involved. There are many types of studies conducted for a wide range of purposes. Depending upon the discipline, groupings that describe types of studies and their definitions vary. Here, we have adopted five of the six groupings referred to by Leek and Peng (2015) as common in bioacoustics. These study types include descriptive, exploratory, inferential, explanatory (called ‘causal’ in Leek and Peng 2015), and predictive studies. Definitions we give here have been framed within the context of common bioacoustical questions, and thus are adapted from more broad definitions.

Of the study types, descriptive studies are the simplest, aiming to summarize datasets collected. Exploratory studies take a step beyond and explore relationships, trends, and patterns in datasets. Neither of these types of studies attempts to infer beyond the dataset collected to the wider population. These types of studies are

commonly used during preliminary data exploration before undertaking inferential, explanatory, or predictive studies (see Sect. 9.3.3). Indeed, descriptive and exploratory surveys are often used to develop the more complex inferential, explanatory, and predictive study type questions. Inferential studies build on descriptive and exploratory studies by quantifying whether findings are likely to be true for a broader population and hence can be generalized. For example, inferential studies are commonly used to make decisions about whether there is sufficient evidence regarding observed patterns or relationships in sample data to believe that they have not arisen from the population by pure chance alone. Explanatory studies aim to identify associated conditions (e.g., species, age, sex of an animal, date, time of day, season, and environmental factors such as temperature, noise, etc.) influencing or explaining an outcome (e.g., the rate at which animals produce their calls). These studies seek to determine the magnitude and direction of relationships (Leek and Peng 2015). Predictive studies aim to predict future outcomes in given conditions or scenarios (but may not necessarily explain conditions leading to an observed outcome). By identifying which of the study types your research aim falls into, the general structure of the analytical question can be formed. Some examples of the different study types and corresponding analytical questions are given in Table 9.1.

9.3 Designing the Study and Collecting Data

Once the analytical question has been formulated based on the study type, novelty, and whether it truly addresses the research question, the feasibility of collecting the required data will need to be assessed. Practical considerations, for instance, include identifying any hindrances to study site accessibility or timely ethics approvals and animal experimentation permits. Below (Fig. 9.1) is a checklist of some preliminary considerations before committing to developing, designing, and executing a study.

Table 9.1 Examples of study types and their corresponding objectives and questions

Study type	Purpose	Example objective	Example questions
Descriptive	Studies conducted to describe phenomena and conditions measured during a study.	Describe the characteristics of sound produced by sea turtle hatchlings recorded during a study.	<ul style="list-style-type: none"> • What is the frequency range of sounds produced? • What are the source levels of sounds produced? • What is the rate of sound production by sea turtle hatchlings?
Exploratory	Studies exploring relationships, trends, and patterns in datasets (not in a broader population).	Establish how observed hatchling sea turtles' sound production varied during a survey.	<ul style="list-style-type: none"> • How does observed hatchling sea turtles' sound production vary during a given survey?
Inferential	Studies aiming to estimate population parameters or test hypotheses about a broader population.	Determine the average expected sound production rate of a population of hatchling sea turtles.	<ul style="list-style-type: none"> • What is the average expected sound production rate of a population of hatchling sea turtles?
Explanatory	Studies that aim to understand the underlying cause(s) of a behavior, state, or phenomenon.	Identify what influences sound production in sea turtle hatchlings.	<ul style="list-style-type: none"> • Are communications influenced by the presence of other sea turtles, environmental conditions, or human/predator threats?
Predictive	Studies that aim to predict an outcome (such as animal behaviors) in response to a stimulus or condition.	Predict hatchling sea turtle sound production rate when threatened by humans.	<ul style="list-style-type: none"> • What will be the expected sound production rate of hatchling sea turtles when exposed to human threats?

<input checked="" type="checkbox"/> Has the question been already answered in past research? <input checked="" type="checkbox"/> Does the analytical question address the research aims? <input checked="" type="checkbox"/> Will there be any logistical / ethical constraints that will affect the execution of the study?
--

Fig. 9.1 Checklist of some considerations to be made before committing to a study

9.3.1 Experimental Design

The ideal situation is to formulate the analytical question before data are collected (i.e., a priori) so that experiments can be designed to maximize the chance that, based on the observations, they produce precise (i.e., close to one another) and accurate (i.e., proximal to true values) estimates of the parameters of interest, and so that there is a high probability of detecting relevant effects (i.e., that there is sufficient statistical power) when they are present. In some cases, however, formulation of the analytical questions occurs after data have been collected (i.e., a posteriori). This may occur as a result of poor planning or of new and unforeseen research opportunities. A scenario in which this often occurs is when data already

collected for another primary study are used to answer a new research question. In these cases, the methods and experiment are not necessarily designed according to the analytical requirements of the new research question. Bioacoustical studies using pre-existing opportunistic data often do so because collecting new data can be prohibitively expensive (e.g., if the field site is remote or if specialized equipment is required). Since the methods and experimental design may be sub-optimal for the current study questions, the data must be meticulously evaluated to check that newly formulated analytical questions can indeed be answered. Studies attempting to answer specific research questions using sub-optimal or poor-quality data cannot always be salvaged, even with sophisticated analyses. The prominent

twentieth century biostatistician, Sir Ronald Fisher, illustrated this problem with the following quote: “To call in the statistician after the experiment is done may be no more than asking him to perform a post-mortem examination: he may be able to say what the experiment died of” (Fisher 1959). This message cannot be overstated. It is critical, wherever possible, to consider the question carefully a priori, so that the study is able to answer the question (Cochran 1977). If you think you might need to consult with a statistician, do so before collecting the data.

For analyses to answer ecological research questions, the experimental design must yield sufficient information about the question of interest. Often, ecological questions involve sets of sampling units taken from a larger group (i.e., the statistical population, hereafter referred to as a population unless otherwise stated). For a given study species, or set of species, sampling units could be defined as individuals, groups, cohorts, communities, or local populations of the species of interest—it depends on the research question. Usually, due to logistical and time constraints, it is not possible nor desirable to make measurements over all objects or the whole population. In these cases, a sample is taken and data collected from the sample are considered to be representative of the population. It is key that the process used to draw the sample is well understood and is ideally random in design. The process of drawing conclusions regarding a population based on a sample from it is called statistical inference.

To make meaningful inferences about the properties of a population, the sampling protocol must yield a sample size that is sufficiently large to represent the population. In addition, the sampling protocol should either eliminate or control significant sources of error including random and systematic error (Cochran 1977; Panzeri et al. 2008). Random error is caused by unknown and unpredictable changes, such as in the environment, in instruments taking measurements, or as a result of the inability of an observer to take the exact same measurement in the same way. Statistical methods typically quantify this error and, in fact, build on it to draw inferences. In some sense,

if there was no error then there would be no need for statistics. Of course, the performance of the analytical methods is affected by the amount of error in the data, in that the statistical power to detect significant effects decreases with increasing error, but if there was no error, by definition there would be no questions left to answer and statistics would have no role to play. Systematic error (i.e., bias) is consistent error that is repeatable if the data are recorded again. It can arise from many causes, such as a person consistently making the same erroneous observation (i.e., biased observation; e.g., incorrectly recording male birds as female birds) or an incorrectly calibrated instrument. In behavioral studies, biases in collected data can also be introduced by the presence of the researchers themselves (e.g., through human disturbance in a study on supposedly undisturbed animal vocal behavior). The introduction of bias can be further illustrated in the example of a bioacoustician estimating acoustic cue production rate (i.e., number of cues, such as calls, produced per unit time) for a population. In this example, the researcher obtains samples of animals by locating the animals producing acoustic cues. It is highly likely, however, that the sample collected will be only from animals that are in a sound-producing state (as silent animals will go undetected), hence acoustic cue rate might be inadvertently overestimated. Furthermore, animals may respond to the presence of the researcher by altering their cue production rates, thereby introducing further error to cue rate estimation. Such studies should be designed to remove or control biases. If controls cannot be integrated into the experimental design, then these may be able to be applied at the analytical stage (statistical controls; see Dytham 2011) and estimation of, and adjustments for, unavoidable biases may be made during the analysis. For topics on experimental design (e.g., systematic, stratified-random, and random-block) that aim to reduce biases and increase inferential power, the reader is referred to textbooks such as Lawson (2014), Manly and Alberto (2014), Cohen (2013), Underwood (1997), and Cochran (1977), among many others. It is critical that researchers carefully consider and identify the

- | |
|--|
| <ul style="list-style-type: none"> <input checked="" type="checkbox"/> Does the scope of the experimental design match those of the questions? <input checked="" type="checkbox"/> Is the sample size large enough given the effect size (see Section 9.5.1.2 for discussion on effect size) being investigated? <input checked="" type="checkbox"/> Are the resources (e.g., time, money, and trained personnel) available for the project sufficient to carry out the study? <input checked="" type="checkbox"/> Will data be reliable (i.e., accurate and precise) enough to answer the questions? <input checked="" type="checkbox"/> Will causes of biases in data collected be able to be identified and removed or addressed adequately? |
|--|

Fig. 9.2 Checklist of some considerations to determine whether a research question can be answered

most suitable sampling design for their research questions.

Despite all attempts to obtain reasonable sample sizes, minimize biases, and carefully select an appropriate experimental design, data quality is frequently sub-optimal due to logistical or practical constraints. Often unexpected restrictive weather conditions and/or failure of instruments limit data collection during fieldwork. Good planning can mitigate unexpected data limitations, thus wherever possible, there should be contingency plans in place to deal with the unexpected (e.g., budgeting for a reasonable number of poor-weather days or redundancy in instrumentation). Even with careful design and contingencies implemented, data limitations can still occur and may need to be dealt with at the analysis stage. However, as noted before, sophisticated analyses to deal with these are always a second-best option over implementing data collection methods and survey design that are robust to potential limitations. Figure 9.2 gives a list of some considerations to be made for assessing whether research questions can be answered before data are collected.

9.3.2 Instruments and Measurements

Instruments must be able to measure subject behavior and conditions of interest in the study such that estimates derived from the observations

have sufficient accuracy and precision to detect the effect(s) of interest. The accuracy of an estimate is its proximity to the true value, while precision refers to the variability of successive estimates of the same quantity. Naturally, to be able to derive accurate and precise estimates, measurements must also be accurate and precise. Accuracy and precision of measurements are evaluated through calibration and testing of the instruments. Some instruments may simply not have the capacity or range required for the study. For example, a low-frequency acoustic recorder will not have the capacity to measure the acoustic behavior of bats, which produce high-frequency echolocation signals. While careful consideration must be made in selecting instrumentation, considerable advances in their capacities have been made over recent decades. Instrumentation in bioacoustical studies is discussed in detail in Chap. 2. Below is a checklist for evaluating whether the selected instrumentation will collect the required data for a project (Fig. 9.3).

9.3.3 Preliminary Data Exploration

Data quality resulting from the experimental design, selected instrumentation, and measurements must be checked through data exploration and visualization (e.g., graphics, spectrograms) before embarking on planned

- Do the instruments have the sensitivity (i.e., sufficiently low noise floor and thus sufficiently low amplitude that can be recorded), dynamic range (i.e., range of amplitudes that can be recorded), frequency range (for sound recorders), and field robustness required for the study?
- Do the instruments obtain sufficiently accurate and precise measures?
- Is there a quality-control process to ensure that instrument accuracy and precision can be measured over time (e.g., systematic calibration and testing)?
- Are the instruments reliable in that they will not result in significant sets of missing or biased data?

Fig. 9.3 Checklist of example considerations for selecting instrumentation for a bioacoustical study

analyses. It can be said that it is never early enough to explore data, nor can there be too many graphs involved in doing so. In fact, a preliminary exploration of data should always be conducted at the beginning of data collection to allow the structure of the data to be investigated, including the presence of anomalous data points, missing values, and potential biases. By identifying these early in the study, unforeseen design, sampling, or instrumentation issues can be rectified. Preliminary exploration of data, after data collection has been completed, will allow for any remaining anomalies and biases to be identified and planned analyses refined. Suspicious observations can be introduced at different stages of the research, for instance through: (1) data entry error, (2) changes in the measurement methods, (3) experimental error, or (4) some unexpected, but real variation. For the first three cases, the anomalous value(s) might be removed before analysis. In the last case, there could be some biologically important reason for the observed unexpected values. Sometimes the word “outlier” is used to refer to these suspicious observations, but we prefer to avoid the term. An outlier implies something that was unexpected, but only after defining what would be expected can we decide what the word “outlier” means. Often “outliers” are very informative and can even lead to new research questions. Consequently, it is important to understand how anomalies have occurred and to ascertain whether they should be removed or not. A good and

honest approach, with little added cost, is to present and discuss the results of an analysis with and without those observations. This approach provides useful information about the practical consequences of the presence of anomalous observations.

If sufficiently large gaps in information from missing values occur, the data may not be representative of the larger population, especially since it might be hard to determine after the survey whether the data were missing at random. Similarly, if measurements were collected under certain conditions (e.g., poor weather or noise), the data cannot typically be used to make inferences outside this range of conditions (which would be referred to as extrapolation). Finally, data of very poor quality may not be salvageable, and—as mentioned before—it is far preferable to get the data right in the first place than to trust analytical solutions to deal with problems introduced at the data collection stage. Data exploration and visualization are further discussed in Sects. 9.4 and 9.5.

9.4 Data Types and Statistical Concepts

Regardless of the analytical approaches used, there are some fundamental terms and concepts that need to be understood before embarking on analyses.

9.4.1 Variable Types and Their Distributions

Measures of observations or conditions of interest in a study can be called variables. For instance, variables can be measurable properties of animals, their behaviors, or their environment. In a study of the acoustic characteristics of elephant vocalizations recorded at different ranges from the animal, relevant variables might include the range between the microphone and the elephant, the subject (i.e., which animal it is), the sound type, the received sound level, the spectral characteristics of the sound at the receiver locations, and the acoustic characteristics of the environment between the elephant and the receiver. In general, a researcher will have a good idea about the plausible values for the variables of interest, and hence what range of values to expect, but not know the exact values before the observations are made. Variables of known expected range but whose exact values are unknown until observed are random variables by definition. The notion of “outlier” is related to this expectation, as “unexpected” values might be considered suspicious. Within a regression context (see Sect. 9.4.3 for more detail), the variables that represent the outcome of interest are called dependent variables or response variables. When they represent the conditions that influence the outcome, they are called independent variables or explanatory variables, sometimes known as predictors or covariates. Hereafter we use all terms to discuss variables, choosing each time the definition we feel will help to make the meaning of a concept most intuitive.

Variables can be of two types: (1) categorical, which can be further subdivided into nominal or ordinal (if there is an order), and (2) numerical, which could be discrete or continuous. Categorical variables are often called factors and are qualitative. For example, if the variable was a sound type produced by a bird categorized as either song or chirp, then sound type would be a nominal factor with two levels, also called a binary variable. If the bird species was known to produce three different sound types, then the

corresponding factor would have three levels. Numerical variables are quantitative, and can be discrete (e.g., integers such as counts) or continuous (where, by definition, an infinite number of values are possible between any two values). Examples of continuous variables are the height and weight of an individual or pressure and temperature, while the number of sounds or the number of individuals are examples of discrete variables. A summary of variable classification and metrics is given in Table 9.2.

Properties of these variables, such as central tendency measures like the mean, mode, and median, or measures of spread like variance and standard deviation, are statistics that can be used to describe a sample of values. When these refer to the values that these quantities have in the population (as distinct from a sample of that population), these properties are called parameters.

Often, additional variables are collected that are not necessarily of interest in explaining a research question but could influence the response variables. For example, while a bioacoustician might be interested in measuring the rate of vocalization of chicks as a function of the parents’ presence, the frequency of predator visitation could also influence vocalization rates. In this example, collecting information on the main independent variable (parent presence) and the variable not of direct interest (predator presence) would be considered important to capture all variables influencing vocalization rate. Some of these variables might be of direct interest, but some might just be included in a study because they can affect the response, and if ignored, would confound the results. For this reason, they might sometimes be referred to as confounding factors or confounding effects. Note that these terms and their definitions vary with discipline (e.g., there is some discussion about the exact definition of a covariate; see Salkind 2010) and analytical software, and sometimes are used interchangeably. Therefore, the reader should make sure that, when reading a source or when reporting their own results, the context provides the required clarity for the wording chosen.

Not only are variables described according to the properties they measure and whether they are

Table 9.2 Variable classification and metrics

	Categorical		Numerical	
	Nominal	Ordinal	Discrete	Continuous
Description	Non-ordered categories	Ordered categories	Variables in which the data can take on only certain values (i.e., values that have non-infinitesimal gaps between them containing no values)	Variables in which the data can take on real values and all infinitesimal real values between them
Example	Sound type (e.g., downsweep, upsweep, constant tone)	Vocal activity on a scale ranging from not vocally active (0) to highly vocally active (5)	Acoustic cue counts	Received sound exposure level (in dB)

independent or dependent variables, but in the context of some analytical methods (e.g., linear regression models and their extensions) they are also described by whether they represent a specific or random set of values. Generally, in statistics, a variable with a value that is not known before it is observed (e.g., peak frequency of a call or number of animals in a group), but of which the range of possible values is known (e.g., a positive continuous number like the amplitude of a lion’s roar), is known as a random variable, as described above. Its range of possible values is referred to as the domain of the random variable.

A random variable can be characterized by its probability distribution, which describes the probability of observing values in a given range of the domain of the variable. An infinite number of distributions exist, but some, given their useful properties, are widely used. These distributions are given names so that we can easily refer to them. Arguably, the most widely used are the Gaussian distribution (perhaps more often known as the normal distribution, but since there is nothing normal about it and it induces practitioners to think there might be, we avoid the term here), gamma distribution, and beta distribution, used to model continuous data; while the Poisson distribution, negative binomial distribution, and binomial distribution are useful when modeling discrete values. The uniform distribution is one in which all values in the domain are equally likely and can be either continuous or discrete. These distributions are typically defined by their parameters. As an example, the normal distribution is defined by the mean and the

standard deviation, and for the case of the Poisson, it is defined by the mean only. Given the parameter values that define a random variable, all the characteristics of the random variable are unambiguously defined.

Values of a discrete variable are characterized by a probability mass function (pmf). A pmf is a function that gives the probability that a single realization of the variable takes on a specific discrete value. The number of vocalizing individuals detected in an area might be approximated by a Poisson random variable, characterized by its mean (such as 3.7 individuals). The Poisson distribution is special in that its variance is equal to its mean, a restriction that means that often it does not fit biological data well, where larger variance than the mean is the norm.

In contrast, continuous variables can be characterized by a probability density function (pdf). In the instance of a variable such as the change in duration of song, the pdf might be represented by a Gaussian distribution—a bell-shaped curve characterized by its mean and standard deviation. For example, the variable “change in song duration” could have a true mean change in duration of 240 s and a true standard deviation of 12 s. These true values are generally unobserved, but we would like to estimate them. A single measurement of change in song duration by a researcher could produce a value of 228 or 271 s. These single values are referred to as realizations of the random variable. Pdf functions provide information about how the values are distributed before they are observed. Further

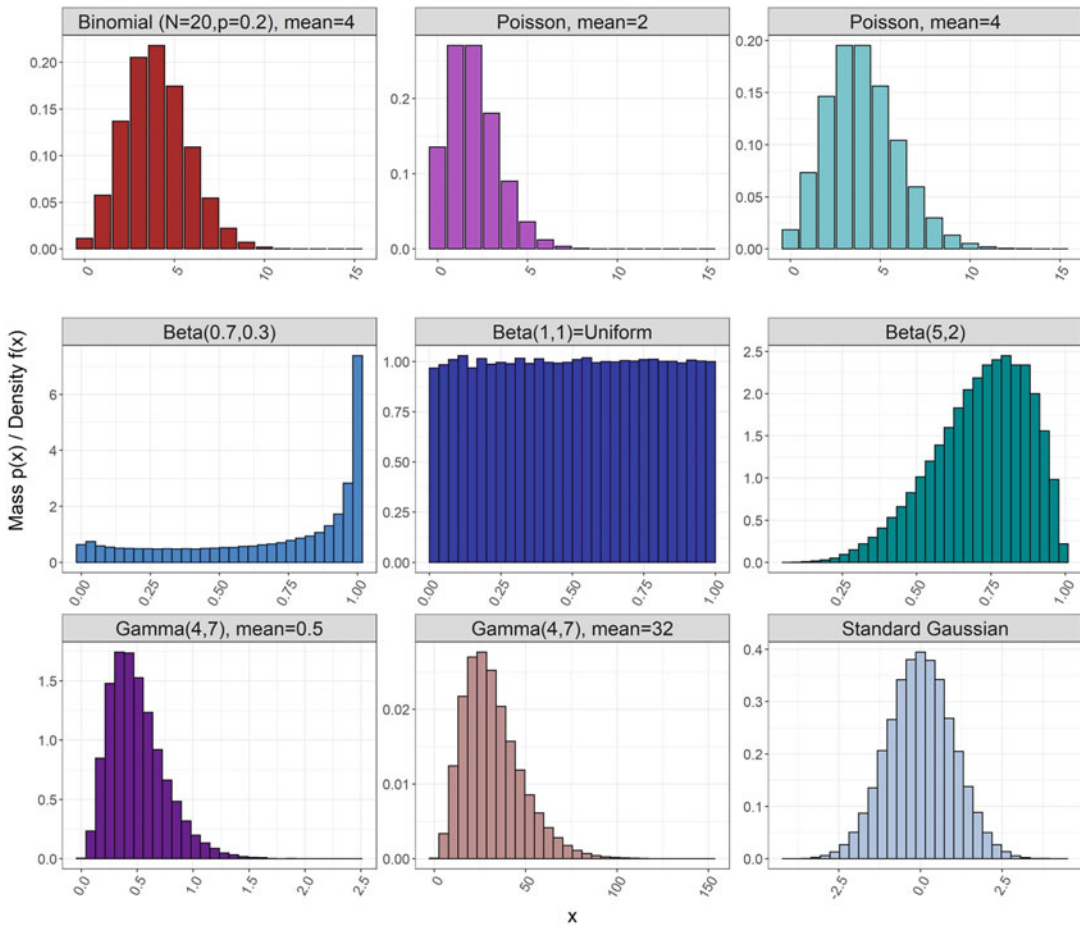


Fig. 9.4 Examples of samples taken from different distributions. The Gaussian, gamma (defined by its shape parameter k and scale parameter θ) and beta (defined by shape parameters α and β) are continuous distributions, represented with histograms. The Poisson (defined by its mean) and binomial (defined by n independent

experiments and outcome success probability p), represented with barplots, are discrete distributions. Note some distributions can be special cases of others. As an example, the beta distribution, with shape parameters $\alpha = 1, \beta = 1$ is shown, illustrating the fact that it is equivalent to a uniform distribution

examples of distributions are given in Fig. 9.4. The reader is referred to Quinn and Keough (2002) for a good introduction to useful probability distributions in biostatistics.

9.4.2 Estimators and Their Variance

In this section, we introduce estimators and related concepts because we will need them later, but we note that we do so very briefly, just

so that the terms do not come as a surprise. The reader is referred to Casella and Berger (2002) for further details on statistical inference, estimators and their variance.

As discussed previously, a parameter is a quantity relating to the population of interest. When performing statistical inference, we want to estimate the parameters in the population (e.g., the mean cue production for a species of whale) using samples (e.g., a sample of acoustic tags put on whales). To estimate parameters, we use

estimators. An estimator is a formula that we can use to compute a parameter based on a sample. In the case of estimating the population mean, the estimator is, not surprisingly, the well-known formula for the sample mean. Estimators are therefore based on random variables, in the sense that each time we collect a sample we would get a new observed value (i.e., a new estimate). Thus, an estimator can also be thought of as a sample statistic that estimates the population parameter such as the mean. If we collected infinite samples and computed the estimator each time, we would get the estimator sampling distribution, from which we could evaluate the bias and the variance of an estimator. However, collecting infinite samples is not possible, but by understanding the properties of the estimator and the design used to collect the data, we can also quantify the variability associated with an estimator, based on a single sample. Variability is a key attribute of an estimator, and the resulting estimate from the single sample (known as the point estimate) is not enough to provide a full representation of it. For example, it is very different to say that we estimate a cue production rate to be 7.2 sounds per hour, than to provide the additional information that it could vary from 7.1 to 7.2, or that it could vary from 1.2 and 27.7. In the first example we have a small variance, and the latter we have such a large variance that the estimator itself is borderline useless. To compute an estimator's variance, there are two main approaches. If the estimator and the process by which we collect the sample is simple enough, we have standard formulae for the variance. That is the case for the sample mean from a simple random sample. However, often in practice, that is not the case, say because the sampling procedure is convoluted, there is a hierarchy in the process, or the estimator is composed of several random components, possibly not independent among themselves. A good example is an animal density estimator from Passive Acoustic Monitoring (PAM), where different random components like encounter rate, detection probability, cue rate, and false-positives might be at play (see Sect. 9.6.2 for a PAM density estimation example). In such cases, resampling techniques like the bootstrap

might be considered. The rationale behind the bootstrap is that one can resample with replacement from the original sample, and the variability of the estimates computed over the resamples is an estimate of the estimator variability. The reader is referred to Manly (2007) for further details about these procedures. While variance is commonly reported, when comparing variances of quantities that have different means, the coefficient of variation (CV), which is the standard deviation divided by the mean, can be useful. The CV is typically reported as a percentage ($\% \text{ CV} = \text{standard deviation}/\text{mean} \times 100$).

9.4.3 Modeling

In its most simplistic form, a model is a mathematical generalization of the relationship among processes (Ford 2000). Models are by necessity a simplification of reality. Extending a quote popularized by George P. Box (1976), all models are strictly wrong, in that they are always oversimplifications of reality, but many models are useful, in that they provide useful explanations or predictions of reality. Models can either be empirical or theoretic. A common example of a theoretical model in acoustics is the piston model used to represent the beam pattern in a directional sound source like the dolphin biosonar system (Zimmer et al. 2005). While theoretical models are based on theory, empirical models are based on observations. Here we will focus discussion on empirical models as observed data are commonly used to fit models to describe bioacoustical processes. Models describing the relationships between whale vocalization rates and season or location (Warren et al. 2017) or dolphin occupancy and pile driving noise (Paiva et al. 2015) are examples of empirical models. Another example is a mathematical equation that describes the number of bird calls recorded within a given period as a function of the number of birds present. By identifying the mathematical relationship between variables, past events can be explained and future scenarios predicted. However, finding such an association requires careful interpretation, especially in observational

studies. Finding an association between two or a set of variables does not necessarily imply a causation. This could be either a spurious association, or an observation induced by a variable that was not recorded. It is a statistical capital sin to confuse correlation with causation. For example, on hot days, the consumption of ice creams increases, and so does the number of fires. But you can eat an ice cream guilt-free as you will not cause a fire!

9.4.3.1 Introduction to Regression: The Cornerstone of Statistical Ecology

Arguably, the most common and most useful class of statistical models are regression models. The simplest regression model (i.e., the Gaussian linear regression model) has three basic components: (1) a dependent variable that is to be modeled (i.e., described or explained), and (2) independent variables that are thought to influence the dependent variable. The third component, the random error, distinguishes statistical models from deterministic mathematical models. The random error captures how the model differs from the actual observations. In other words, it measures how well, or badly, our model describes reality. Written as a mathematical expression, the simple regression model looks like this:

$$Y = \alpha + X\beta + \varepsilon, \quad (9.1)$$

where Y is the response variable, α is the intercept (a constant), X is the fixed independent variable, β is the regression coefficient for the fixed independent variable that describes the rate of change of the response variable as a function of the independent variable, and ε is the random error. In general, the parameters α and β are not known and must be estimated based on data.

Most variables, particularly in ecology, are influenced by many covariates, and hence models can include multiple independent variables. For instance, in a study on whether the vocalization rate of sea lions differs with sex and age, vocalization rate (i.e., number of vocalizations per unit time) would be the response (dependent) variable and sex and age the explanatory (independent)

variables. In addition to having these two explanatory variables of direct interest, other variables may also be relevant to include in models, because they might a priori be expected to also influence the response variable. Variables that may affect vocalization rate may include time, season, social context, or location. Studies in which multiple explanatory variables influence the outcome might have interactions between the explanatory variables that are important to consider. For instance, vocalization rate may differ between male and female sea lions, but only for sub-adults and adults and not for pups and juveniles.

In a regression model, a distribution is typically assumed for the response variable. This will induce a distribution for the random errors. Historically, regression models considered the errors of the dependent variable to be Gaussian distributed, and much of regression theory was developed under this assumption. Note that a model assuming a Gaussian error distribution in the dependent variable is commonly simply referred to as a linear model. Nowadays many generalizations to linear models exist (as described below and see Zuur et al. 2009 for common examples in ecology; see Generalized Linear Models in Sect. 9.5.3 below). Arguably, as noted above for random variables, the more commonly used distributions in regression models are Gaussian and gamma for continuous data, Poisson and negative binomial for counts, binomial for binary data, and beta for proportions (or probabilities), but many others exist. As for linear models, generalizations assuming other distributions associated with the response variable and associated error structure are commonly referred to by their distributions. For example, a Poisson distributed response variable with associated error structure of counts of animals is commonly referred to simply as a Poisson model. A gamma model might be used to model continuous positive values resulting from measurements of duration of a recorded song. Values representing the probability of producing a sound (between 0 and 1), however, might be modeled assuming a beta distribution.

Regardless of the error distribution of a model, classical regression models assume that observations are independent of each other (i.e., the value that one observation takes on is not influenced by another). The easiest way to ensure this happens is by design, and all efforts should be made to enforce it. In the biological world, the assumption is very often violated, and almost as often ignored. This can lead to errors in inferences made, the severity of which depends upon the degree and type of non-independence between observations. A few obvious sources of lack of independence (i.e., dependency) are observations collected within groups that share a characteristic (e.g., a litter or a pod of animals), or observations collected over space (where two observations are more likely to be similar the closer they are in space) and over time (where two successive observations are more likely to be less independent than two observations separated by a longer period of time). Researchers often mistakenly analyze data collected without proper consideration of whether observations are independent. By exploring and accounting for dependencies, or even purposefully including them in an experimental design, the power of an analysis may be enhanced. As an example, in a repeated measures study of bird vocalization rate as a function of time of day, repeated measurements of the same individuals during the day and night could be undertaken by design (instead of randomly sampling birds at each time period). Another example is that of a chorusing group of insects, in which sounds can be produced for hours. A researcher may be interested in measuring whether the insects chorus in a given 5-min period. At any point of time within a chorusing bout, the probability that insects will be chorusing in a 5-min time window will be expected to be high if they were chorusing during the previous 5 min. This leads to what are called autocorrelated observations. In such cases, the autocorrelation structure can be incorporated into the model. If evaluating the effect of time was not of specific interest in this study, an alternative and simpler solution would be for the model to use subsampled data to include only times at which insect sound production can be considered

independent. However, by explicitly accounting for the autocorrelation structure in the model, more efficient inferences are bound to be obtained as there is no loss of information. Model implementation does become a bit more complex, however. Studies that purposefully measure subjects or populations repeatedly over time to create a time series of data are called longitudinal studies. Because time-series measurements, such as those from longitudinal studies, usually cannot be considered independent from one another (e.g., an animal's current behavior is likely dependent on its behavior during the previous sample time), a wide range of models have been purposefully developed to account for non-independence (see Sect. 9.5.3). Researchers should carefully consider and plan for potential sources of dependency in the design of their studies and data collection protocols.

A checklist of some considerations for describing and defining variables in your study, including whether they are autocorrelated or not, is illustrated in Fig. 9.5. These considerations should be made as part of the experimental design and analytical planning process prior to data collection and will need to be reassessed post data collection.

9.5 Tackling Analyses

In this section, common analytical approaches used in descriptive and exploratory studies are presented first, followed by those used in inferential, explanatory, and predictive studies. It is important to note that analyses relevant to inferential, explanatory, and predictive questions require preliminary data exploration (see Sect. 9.3.3), thus requiring descriptive and exploratory analyses first. In these cases, preliminary exploration of data attributes may refine previously planned analytical approaches. This is particularly relevant since sufficient data quality and specific distributions are required for empirical model assumptions to be met and these features can be assessed via initial data exploration.

Analytical approaches described in this section are examples only of a wider range available. The

- Have the variables and variable types been identified?
- If there are multiple independent variables, are there interactions of interest and/or are any of variables highly correlated?
- Are data for variables likely to be independent or autocorrelated?

Fig. 9.5 Checklist of some considerations for defining variables in your study

purpose is, by way of examples, to provide a taste of the explosion of tools developed over the past few decades, the lively discussion that has arisen from their varied and inherent limitations, and the resulting developments in statistical approaches. The reader is directed to the wide range of available statistical textbooks and scientific papers to gain an in-depth understanding of the full range of approaches, their underlying concepts, and their correct use, limitations, and interpretation of outputs.

9.5.1 Descriptive and Exploratory Research Questions

Having defined the question (Sect. 9.2) and identified the variable types and some of their attributes (Sect. 9.4), tackling the analyses is the natural next step. For descriptive and exploratory questions and preliminary data exploration, summary statistics and graphical visualizations provide information about the attributes of variable measures and patterns and relationships in data. The information relates only to the properties of the observed data. Analyses that aim to generalize a sample to a population require inferential, explanatory, and predictive type analyses (discussed in Sects. 9.5.2 and 9.5.3).

9.5.1.1 Univariate Summary Statistics and Graphical Visualization

Exploration and visualization in their simplest forms are undertaken by evaluating each variable on its own (Fig. 9.6). Analyses of single variables are called univariate analyses and are used for representing and summarizing the characteristics of the variable in question. For example, univariate exploratory statistics describe a variable's

properties such as statistics for central tendency including the mean (note that there are different types of means; e.g., arithmetic, geometric, and harmonic), median, or mode, and spread of data including the range (maximum and minimum), variance, standard deviation, skewness (degree of asymmetry), kurtosis (i.e., how peaked a distribution is), or interquartile range (see Table 9.3). Data corresponding to a single variable can be summarized and explored using a range of graphing tools, such as histograms, box plots, bar charts, or scatterplots. Additionally, geographical data can be explored on maps and marine charts, and acoustic spectral characteristics on spectrograms (representing signal strength over different frequencies over time). As noted previously, it is (arguably) almost impossible to produce too many graphs at an exploratory stage—the more that you can learn about your data, the better. The reader is referred to standard statistical textbooks for information on the large range of summary statistics and graphical visualizations available (e.g., Zuur et al. 2007; Zuur 2015; Rahlf 2019 for examples in R).

9.5.1.2 Bivariate and Multivariate Descriptive Statistics

The analyses of two variables together are called bivariate analyses. For instance, exploration and visualization of a given variable as a function of another variable to investigate possible correlation is a bivariate analysis (see Fig. 9.7). A practical example of a bivariate visualization is the use of box plots to visualize the distribution of call types (one variable) as a function of age class (a second variable), or a scatterplot of a recorded acoustic cue rate as a function of time of day. Following this logic, multivariate analyses

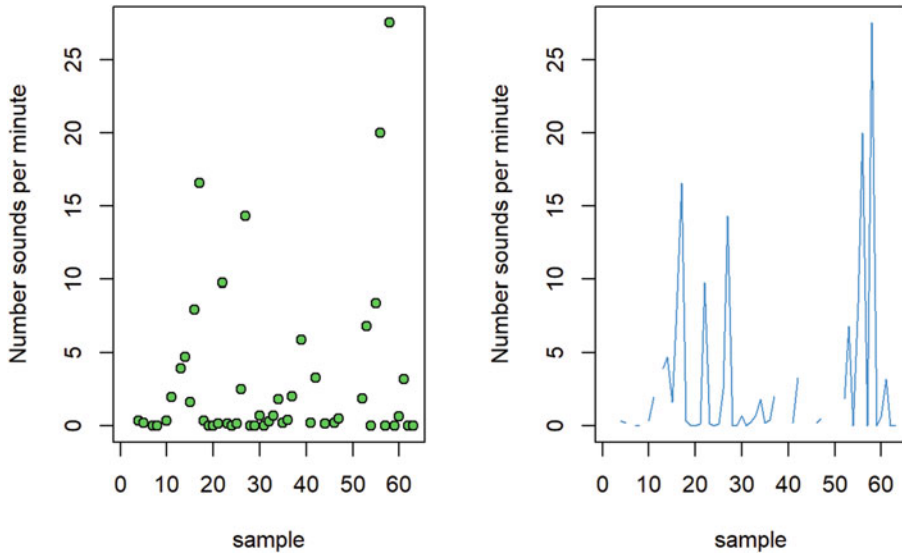


Fig. 9.6 Example of univariate data visualizations of dolphin sounds detected: (left) scatterplot and (right) line chart. Data source: WAMSI as part of Project 1.2.4 (Brown et al. 2017)

Table 9.3 Description of example univariate analytical and visualization tools

Measure	Statistic	Visualization tools	Common purposes
Location and central tendency	Mean (arithmetic, geometric, harmonic), median, mode	Point, line and bar charts, histogram, boxplot	Describe the central tendency of values in a variable
Spread	Range (maximum and minimum), variance, standard deviation, skewness, kurtosis, standard error, interquartile range	Scatter plots, box plots, interquartile range, point, line and bar charts with standard error bars	Describe the spread of measures in a variable and identify patterns and data gaps

naturally consist of the joint analysis of multiple variables. Visualization tools and summary statistics can also be applied to multivariate analyses. For instance, two and three-dimensional scatterplots, bar charts, stacked bar charts, and multiple line graphs can display statistics and spread of data as a function of multiple variables on the same figure.

When bi- or multivariate analyses aim to explore associations and patterns, the magnitude of the association can sometimes be quantified. For example, in a bivariate analysis, the magnitude of the linear relationship between two variables can be quantified using a statistic called Pearson’s correlation coefficient (r). The magnitude of an association such as this one is often

referred to as an effect size. For example, Pearson’s correlation coefficient is a standardized metric ranging from -1 to 1 ; with a perfect negative association yielding a value of -1 , no association 0 , and a perfect positive association a value of 1 . In some disciplines, conventional criteria have been suggested to classify effects as small, medium, and large (see Cohen 1988). What may be in one study considered a large effect (say, $r = >0.6$), however, may not necessarily be in another study (where say, $r = >0.8$ might be considered large). Consequently, evaluating what is a meaningful effect size that a study aims to detect should always guide the design of a study and interpretation of its outcomes. It is a question that the researcher

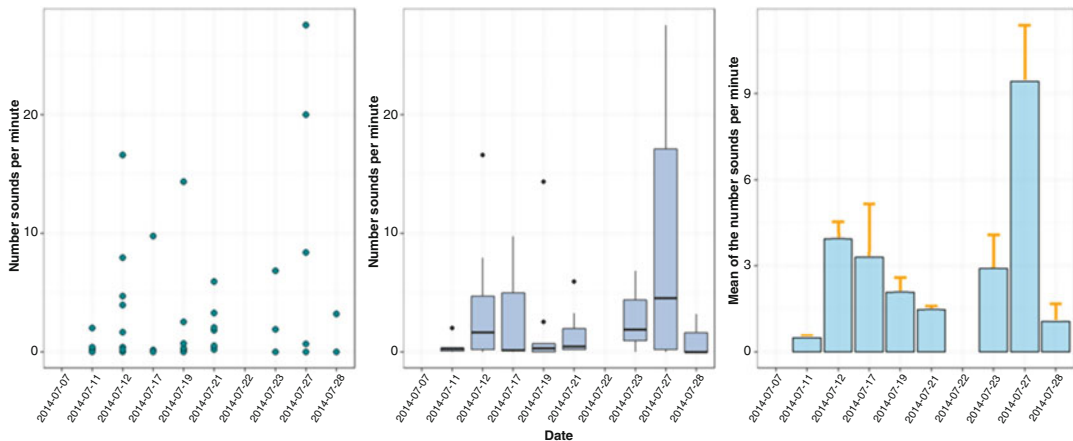


Fig. 9.7 Example of bivariate data visualizations of dolphin sounds detected during July 2014: (left) scatterplot, (middle) box plot, and (right) bar chart with standard error

bars. Data source: WAMSI as part of Project 1.2.4 (Brown et al. 2017)

should answer based on their biological knowledge and is not related to statistical considerations.

When a study's goal is to explore associations and patterns among many variables, analyses become more complex. Multivariate approaches are commonly used to reduce many variables to a few key ones. This is known as dimension reduction. Multivariate approaches are also used to explore relationships and clustering, and to classify objects based on common multiple variable attributes. A good source for additional details on multivariate methods is Borcard et al. (2011).

One of the most common analyses used for dimension reduction is principal components analysis (PCA). The name of the method is derived from the fact that new variables, known as principal components, are obtained from the set of original variables. For example, a researcher may be interested in exploring whether populations of a social insect, such as a species of ant, can be determined based solely on acoustic signals (e.g., stridulations) its individuals produce for communication. In this case, a range of variables might be measured, such as pulse duration, bandwidth, minimum and maximum frequency, and intensity, to name a few. In acoustics, a large number of variables might be measured to capture the full range of

characteristics of acoustic signals. Consequently, using a data reduction method to capture the most variance explained by these variables by creating just one or two new variables (called principal components in PCA) makes the exploration of patterns in sound characteristics easier. The first principal component retains most of the original variance, followed by the second component, and so forth. These principal components are sometimes called factors. Factor 1 and 2 can be plotted against each other, and distinct groupings of plotted values for different populations would be suggestive of differing characteristics in stridulations among populations. To statistically test differences, PCA might be used to generate factor scores as inputs into inferential, explanatory, and predictive analyses (e.g., a regression analysis). Note that there are many dimensionality reduction approaches (see Van der Maaten et al. 2007), and researchers planning on using these tools should acquaint themselves with the wide range available today, their conditions of use, and their limitations. While one approach may be suitable given the attributes of one dataset, another may be required for a different dataset.

Clustering and classification analyses assign objects into groups based on measured attributes (variables). Cluster analyses form groups (McGarigal et al. 2000; Zuur et al. 2009) using

“unsupervised learning,” where you do not “train” the procedure by labeling “training” data with group membership as you might in other methods. A range of cluster analysis algorithms are available including common approaches such as k-means and hierarchical clustering (see Borcard et al. 2011). Clustering and classification are used commonly for pattern recognition and are described further in Chap. 8.

Many other multivariate analytical approaches are available, ranging in their assumptions, strengths, and limitations, and the variable attributes for which they are most suitable. For example, correspondence analysis (CA) is similar to PCA, but can better cope with categorical data. The reader is referred to the many textbooks on the subject, such as Everitt and Hothorn (2011) on some of the more commonly used multivariate methods and their practical application in the software R.

As in the univariate case, we reiterate that associations identified in exploratory multivariate analyses do not indicate causation. Researchers interpreting exploratory analysis results should take care to never conclude that the results are evidence of causation. A brief checklist has been provided below with examples of the types of data considerations required for selecting analyses suitable for descriptive or exploratory questions (Fig. 9.8). The checklist is not exhaustive, rather it is indicative of the kinds of considerations required.

9.5.2 Inferential Studies

Statistical inference is used to infer properties of a population (e.g., estimate parameters) or test hypotheses. There are two widely used distinct frameworks for making statistical inferences: the frequentist and the Bayesian paradigms. Classical frequentist inference has a long history and has dominated past animal behavior and ecology research, while Bayesian inference is becoming increasingly popular. Both approaches can provide insightful information, however, they represent different interpretations of probability.

In frequentist probability, the probability of an outcome occurring is based on the relative frequency of occurrence based on a large number of observations taken. For example, the probability of bird vocalizations being recorded at a study site might be based on many sample recordings taken under the same conditions at the site. If vocalizations occurred 48% of the time, the probability of the outcome of birds vocalizing would be interpreted as 0.48. As the sample size increases, the proportion of occurrences approaches the true (unknown) proportion. If the sample size is small, the calculated proportion may not be a reliable representation of the true probability.

In the Bayesian interpretation, the probability is the degree of *belief of the likelihood* of the outcome. For example, it may be that a researcher believed that vocalization in nesting birds is related to predator presence. The researcher had visited the site and rarely heard birds vocalizing when predators were absent but noticed them vocalizing more often when predators were present. Maybe the researcher had even made a few recordings when predators were present and absent and found that birds were vocalizing 5 out of the 10 times she recorded in the presence of predators and 1 out of 10 times in their absence. In this example, these observations would constitute the prior belief. The research then undertakes a study designed for the purpose of collecting an unbiased set of observations to be used in analyses (sampling in the presence and absence of predators). Using Bayes’ Theorem, the prior knowledge can be used to calculate the probability of vocalization that accounts for knowledge before and after collecting evidence (sampling). If the number of samples is large, the resulting probability estimate may not change much from that obtained in a frequentist framework. However, if the sample size is small, the prior knowledge may significantly affect the estimate of probability. Therefore, the lower the sample size (i.e., in general the lower amount of data coming from the data), the more the prior becomes important.

Many professional statisticians fall firmly in the frequentist or Bayesian camp. This often

Do I require description, exploration, and visualization of individual variables, either for answering the main study question or for checking the quality of the data and assumptions of analyses planned for inferential, explanatory, or predictive studies?

The answer to this is always **YES**. Data always need to be checked for quality and attributes, and if the question requires inference or empirical models, the validity of assumptions needs to be checked (see Section 9.3.3 and 9.4)!

What types of variables do I have?

Does the study question involve single or multiple variables?

If multiple variables,

are there a large number of variables that I need to reduce, explore their association, or investigate clustering or classification of groups characterised by them?

Fig. 9.8 Checklist of some considerations for identifying approaches for descriptive and exploratory questions

follows directly from their training, or just by convenience and actually not having thought much about the philosophical ramifications of their choice. Sometimes they are rather inflexible in their beliefs (be it in one or the other camp). We recommend a more pragmatic approach in practice. Depending upon the problem at hand, one or the other framework might be more suited to the question, easier to implement, or more sensible for incorporating all available information (Nuzzo 2014; Ortega and Navarrete 2017). Consequently, we believe that the modern bioacoustician should have a basic understanding of the differences between frequentist and Bayesian approaches, and suggest that rather than only being frequentist or Bayesian, a pragmatic approach be taken. Below, we provide a very brief introduction to statistical inference applied to parameter estimation and hypothesis testing.

9.5.2.1 Parameter Estimation

There are a range of approaches to estimate population parameters, such as the population mean or variance, or a shape or scale parameter of a distribution, from a sample. In the context of ecological modeling, the frequentist approach to estimating parameters typically uses maximum-likelihood (Hilborn and Mangel 1997). In Maximum Likelihood Estimation (MLE), parameter

values of a distribution are estimated by maximizing the likelihood function so that the MLE estimates are the values of the parameters that are most likely given the sample data. An alternative method is Least-Squares Estimation (LSE), where a solution that minimizes the sum of the squares of the residuals (the difference between the observed values and those obtained using the fitted model) is obtained. For a Gaussian-distributed response variable, and several other simple examples, the LSE solution is equivalent to the MLE. Nowadays LSE are mostly introduced for teaching purposes, and most implementations use maximum likelihood.

As indicated above, the Bayesian framework combines information on the likelihood of an outcome using observed data with prior information on the distribution of the unknown parameter being estimated. The prior distribution can be an assumption based on the researcher's understanding and experience of the parameter before the study began or it can be based on the results from a pilot or previous study. Often the prior distribution simply reflects a lack of knowledge and may be uniform over all the possible values the parameter of interest might take (i.e., the parameter space). A posterior distribution (i.e., updated understanding) is attained by multiplying the prior distribution function with the likelihood

function and scaling the result to provide a probability distribution function. All the inferences are then based on this posterior distribution. The posterior distribution thus can be seen as a compromise between the prior information and the information contained in the data, expressed via the likelihood function. There are various resources available for further reading on the Bayesian framework. Ellison (2004) provides an excellent and gentle introduction to the use of Bayesian methods in ecology, while McCarthy (2007) provides a more thorough overview. Stauffer (2007) gives an in-depth introduction to Bayesian and frequentist statistical research methods and Gelman et al. (2013) discuss Bayesian data analysis. *Statistical Rethinking* by McElreath (2020) is a comprehensive treatment for a reader wanting to become fully versed in the Bayesian philosophy, including R code to explore all the key concepts.

When inferential methods, such as those introduced above, are used to estimate parameters from sample data, the inferences we draw from them are uncertain. Confidence intervals (CIs; a frequentist approach) and credible intervals (CrIs; Bayesian counterparts) are tools for expressing our uncertainty about parameter estimates. Confidence intervals, although more widely used, are arguably more difficult to interpret than credible intervals. Confidence intervals give information based on our sample estimate, and by definition, if we repeated the procedure many times, 95% would include the true parameter value. Note a 95% CI does not mean that 95% of the observations lie within the interval, nor that the probability of the true value of the parameter being in the estimated interval is 0.95. After you estimate the confidence interval, the true parameter value either is, or is not, in the interval, even if we do not know which it is. In contrast, 95% CrIs would represent a range of values for which there is a 0.95 probability that the parameter falls in that range. Ironically, what this means is that while most people use frequentist confidence intervals, they often interpret them, incorrectly, as credible intervals. Although credible intervals are intuitively easier to understand, they can be more difficult to calculate than confidence intervals.

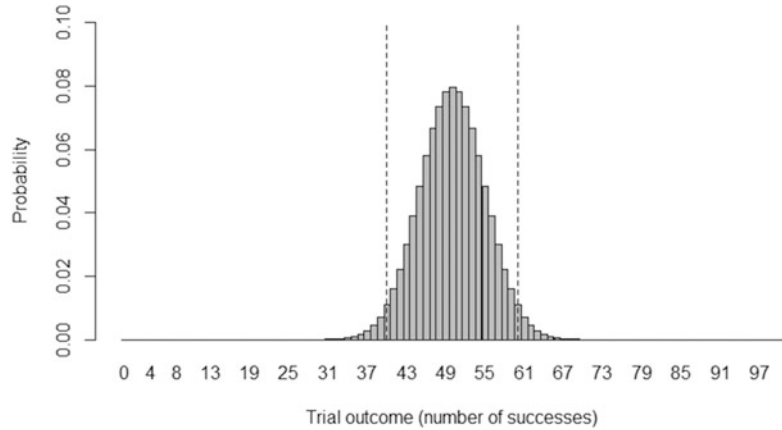
9.5.2.2 Hypothesis Testing

While hypothesis testing has been traditionally undertaken using a frequentist approach (called null hypothesis significance testing, NHST), equivalent Bayesian approaches are increasingly applied. This section focuses on providing a brief introduction to NHST as a foundation and provides references for further reading on Bayesian approaches. These basic concepts are introduced here with examples of their application to test statistics (i.e., statistics values used to reject or support a null hypothesis), however, they are also an integral part of modeling and model selection in explanatory and predictive questions (discussed in Sect. 9.5.3).

NHST constitutes a widespread paradigm under which research has been conducted (NHST, Fisher 1959), however, it is often not used sensibly, and frequently blindly used and abused. In some of these cases, pressure on researchers to find statistically significant effects has resulted in poor research practices (see Nuzzo 2014; Beninger et al. 2012 for detailed discussions on the topic). Applying NHST to reasonable hypotheses and qualifying results according to the limitations and assumptions of NHST, however, can produce important new knowledge. To achieve this, an understanding of how NHST works is required. Here we provide insight into the framework by way of example.

Under the NHST framework, researchers put forward a hypothesis (i.e., proposed explanation) about the phenomena being studied based on a study question. Let us say the researchers' question is "Do seal pup call rates differ between night and day?" The null hypothesis (H_0) is that call rates do not differ between night and day, and the corresponding alternative hypothesis (H_A) is that pup call rates do differ between night and day. Note that this hypothesis implies a two-tailed test, one for which the null hypothesis is rejected if a positive or a negative effect (i.e., a large or small value of the test statistic) is found. In contrast, a one-tailed test would be used by a researcher interested only in the difference between groups in a specific direction (e.g., "Are call rates greater during the day than at night?").

Fig. 9.9 Binomial probability mass function with parameters $n = 100$ trials and $p = 0.5$, with the quantiles 2.5% and 97.5% represented by vertical dashed lines. Under H_0 only 5% of the observations would be more extreme than those quantile values



In this example, the researchers cannot measure the call rates of all animals in the population, so they collect a random sample, say of 100 animals. Sampling at random is key to collecting data that represent the broad population, thereby avoiding biases in the parameter estimates. In this example, on a given day, for each animal, the researchers record the number of calls produced during daylight hours and during the night. Let us call the event, in which for a given animal there are more calls during the day than at night, a “success.” If we assume animals operate independently, then the number of successes in the 100 animals provides information about the null hypothesis: the further from the expected number if there were no differences between night and day, the larger the evidence against H_0 . We also assume that the probability of a success is constant and independent across trials and animals. Under H_0 we assume the probability of a success is $p = 0.5$. Under H_0 , the number of successes has a binomial distribution with parameters n (the sample size) and p . The corresponding probability mass function with $n = 100$ and $p = 0.5$ is illustrated in Fig. 9.9.

To test the null hypothesis, the researchers use the number of successes as a test statistic. The test statistic has information about the null hypothesis, and under the null hypothesis, we know the distribution of the test statistic. If call rates are on average the same during the night and day (i.e., H_0 is true), then we would expect that animals have a probability of 0.5 of producing more calls

during the day than at night, and on average T (number of successes) would equal 50 ($T = 50$).

Now imagine that the researchers observe $T = 46$. From Fig. 9.9, $T = 46$ is consistent with the null hypothesis, which we would not reject for the usual levels of statistical significance (see below for a more in-depth discussion of significance levels). On the contrary, consider the case of $T = 11$. This result would have been extremely unlikely under the null hypothesis, and we would be tempted to reject the null hypothesis, implying that differences between night and day might occur.

The example given here illustrates the rationale under NHST, the steps of which are: (1) define the hypothesis, (2) collect the data, (3) calculate a test statistic, with known distribution under H_0 , (4) evaluate how likely (or unlikely) the data would be under the null hypothesis, and (5) if very unlikely, then reject the null hypothesis, but if not unlikely, do not reject it. Consequently, the trick is to put forward a null hypothesis under which the distribution of the test statistic can be evaluated to assess how likely the data are under the null hypothesis. Given the sampling uncertainty (i.e., not observing the entire population), we can make mistakes when making decisions about whether to reject the null hypothesis or not. The confusion matrix in Table 9.4 illustrates the possible outcomes of a decision.

The two wrong decisions we can make are to reject the null hypothesis when it is in fact true or

Table 9.4 Confusion matrix showing the possible outcomes of a null hypothesis decision: correct decisions and Type I and Type II errors. Statistical tests usually

require a significance level (i.e., Type I error rate), which defines the probability of being wrong if the null hypothesis is true

		Decision on null hypothesis	
		Do not reject	Reject
Reality	H ₀ true	Correct decision	Type I error
	H ₀ false	Type II error	Correct decision

to not reject it when it is false. The former is known as a Type I error (i.e., an incorrect rejection, sometimes referred to as a false-positive) and the latter a Type II error (i.e., failing to find a real effect, sometimes referred to as a false-negative). In general, it is believed that Type I error is what we should guard against, with the logic illustrated here as analogous to the legal system: *It is better to have a guilty defendant not convicted than to have an innocent defendant sent to death.* We note, however, that depending on the problem at hand, a Type II error could have a greater consequence than a Type I error. To illustrate this, imagine that you are testing whether the size of a population has decreased below a critical threshold that requires an action for it to not go extinct. If you do not reject the null hypothesis (i.e., that the population size has not changed) but it is false, you might miss the opportunity to take action and prevent the population’s extinction. Alternatively, if you mistakenly take action to protect the population while it is in fact above the minimum threshold, you might waste money but any risk of detrimental population consequences is eliminated. So, while many textbooks may allude to the importance of safeguarding against Type I error, the error type that should be of most concern is likely to be study-specific. The usual advice applies: Do not use cookbook recipes, rather think about your study. The allowable Type I error can typically be specified with a critical significance level value (defined below). Estimation of Type II errors typically requires another step, called a power analysis (see Ellis 2010 for a textbook on power analyses).

In practice, the amount of evidence against the null hypothesis required in a study is given by setting a threshold based on how unlikely the observed data would have to be under the null

hypothesis before it is rejected. Alternatively, we can compute the probability of, given the null hypothesis is true, observing a value for the test statistic that is as or even more extreme than the observed value. This probability value is commonly referred to as the *p*-value. In the above example, assuming a two-tailed test, the *p*-value associated with $T = 46$ or $T = 11$ would be 0.484 and ~ 0 , respectively. This would lead us not to reject the null hypothesis in the first case, but to reject it in the second case. Note that a common error is to confuse the *p*-value with the probability of the null hypothesis being true or the alternative being false. Researchers should take care in their interpretation of *p*-values to ensure they are accurate.

The predefined probability threshold below which we are willing to reject the null hypothesis is called the significance level (typically designated as α). A typical value for the significance level is 5%, with tests having *p*-values lower than 0.05 often being reported as statistically significant. This value has become widely used; however, it should be noted explicitly that there is nothing special about a 5% significance level. While using this threshold has been extremely useful in practice, there is arguably no other concept in statistics that has received more criticism. The abuse of the 5% significance level by blindly using it is among the most common criticisms of the *p*-value and hypothesis testing (Nuzzo 2014; Yoccoz 1991; Beninger et al. 2012). Using common sense is fundamental in selecting significance levels. It is intuitively sensible that it cannot be sound science to blindly claim a result to be significant if $p = 0.049$ but not significant if $p = 0.051$. Ultimately, researchers need to think carefully about the cost of errors they can incur and define suitable significance levels accordingly. The focus should arguably

be on reporting confidence intervals and assessing the biological importance of reported effects, not on claims of statistical significance that are often not more than statements about sample size. Given a large enough sample size, even the smallest difference will become statistically significant. Therefore, it is perhaps not surprising that a common pitfall for researchers, and equally as or arguably more important than evaluating statistical significance, is failure to consider a result's biological significance. Imagine two populations of a whale species that produce the same stereotyped calls. Let us say animals in population A produced calls at a mean rate of 22.7 per hour and in population B at 22.6 calls per hour, and that these are significantly different statistically. Is this result meaningful biologically? In other words, is the effect size of a magnitude that we care about? In most cases, almost certainly not. Therefore, a researcher should have a good understanding a priori of the magnitude of the effect that is biologically relevant. Researchers undertaking studies with large sample sizes having the power to detect very small effect sizes can fall into the trap of reporting results as important based on statistical significance instead of on effect size and significance together. Conversely, studies having a large probability of incurring Type II errors (also known as low power, i.e., having a low probability of correctly rejecting the null hypothesis when it is false) due to a small sample size may only be able to detect very large effect sizes and miss smaller ones that are biologically important. The effect size that is meaningful in a study, thus, needs to inform the experimental design to ensure a sufficiently large sample is collected before the study commences.

While NHST and p -values can provide valuable tools to bioacousticians, it is not amiss for researchers to be well aware of the lively discussion on their misuse, drawbacks, and limitations. Nuzzo (2014) provides an introduction to this discussion, Yoccoz (1991) provides a classical critical review regarding their use in biology and ecology, and Beninger et al. (2012) frame the problem in the wider context of statistics in

(marine) ecology. An entire *Forum* section in the journal *Ecology* has been dedicated to the topic in recent years, and Ellison et al. (2014) show that while having been discussed and revisited many times in recent years, the discussion about their use is alive and kicking!

Having said this, a wide range of NHSTs have been developed over many decades to accommodate a range of questions and data types. Traditionally, many of these have been described as either "parametric tests" or "non-parametric tests," with parametric tests often assuming samples arise from Gaussian distributions and non-parametric tests are often used for categorical or continuous data that do not fit assumptions of parametric tests. While we urge the reader to be cautious about blindly using such tests and be aware of their limitations, we feel we must discuss them since this is how statistics is presented in most undergraduate and postgraduate courses aimed at the applied sciences, biology and ecology included. As examples, tests commonly referred to as parametric include the z -test (for testing a sample mean), t -test (for comparing the means of two groups), and analysis of variance or ANOVA (used for comparing two or more groups). Common non-parametric alternatives to the t -test and the (one-way) ANOVA are the Mann–Whitney U and Kruskal–Wallis tests, respectively. The tests referred to here are only a few of the vast range available, and readers will not find it difficult to find a plethora of textbooks describing them. Note that these tests have been used widely in past decades and continue to be used in current research. Today, however, with improved knowledge of limitations of these tests, they are losing their appeal (see e.g., Touchon and McCoy 2016). In general, they are no longer the standard go-to for particular types of problems as they have been superseded by more robust approaches. With advances in statistics, a wide range of readily available modeling approaches has been developed that more than accommodate data that would have traditionally been analyzed using non-parametric tests (see Sect. 9.5.3 for an overview). Note that while many disciplines are guided by traditional "parametric" and "non-

parametric” classifications, where parametric would often be associated exclusively with the Gaussian distribution, modern approaches in statistical ecology using regression models are generally not said to be parametric or non-parametric; rather, they tend to be referred to based on the data distributions for which they are suited, such as a Poisson or gamma regression (see below for more on these).

9.5.3 Explanatory and Predictive Research Questions

Explanatory and predictive studies have questions requiring a response variable to be described as a function of a set of independent variables. Arguably, the majority of the models used by ecologists to answer this type of question are some kind of regression model. However, these models come in many forms. This section aims to introduce the reader to different types of regression models. We note upfront that model selection and validation, and inference from selected models, are fundamental aspects of these analyses and are only very briefly mentioned in Sect. 9.5.3.1. Relevant yet accessible books with plenty of practical examples addressing these steps include Zuur et al. (2007) and Zuur et al. (2009).

Historically, linear regression models (in which the errors are assumed to follow a Gaussian distribution) were the only tools available to answer this type of question. When the only tool you have is a hammer, all your problems begin to look like nails. With a Gaussian error distribution assumption, the only analytical options are simple linear regression models of the type given in Eq. (9.1) or linear regression models with several predictors (i.e., multiple regression). There are many special cases of such linear normal regression models including the independent sample t-test, ANOVA (i.e., analysis of variance for multiple sample mean comparison), ANCOVA (i.e., analysis of covariance for regressing a continuous response variable on a factor and a continuous covariate), and MANOVA or MANCOVA (i.e., multivariate

extensions of the former methods). Note that these approaches have additional assumptions, such as that of homogeneity of variances. Homogeneity of variance means that the variance for a response variable is assumed to be constant across values of the independent variable. Many datasets have been forced through these methods even when they were clearly not the right tool for the job. This included, for example, transforming the response variable (e.g., by applying a log function to it) until Gaussian distributional assumptions were met to a reasonable extent. But even then, often a method’s assumptions were not met. For instance, there is no transformation that will turn a discrete count into a continuous variable. For an interesting presentation about why not to log-transform data, see O’Hara and Kotze (2010). Nonetheless, sometimes processes might have properties that make a log-transformation of the data sensible and useful (e.g., Kerkhoff and Enquist 2009). While transforming data to fulfill methods’ assumptions has been acceptable in the past given a lack of accessible alternative methods, this is often no longer the case, and successful ecologists need to have a few additional tools in their toolbox. The rule is one that practitioners do not enjoy: There is not a single rule that fits all questions and problems, we need to understand the problem to know how to model it. Sometimes it is even said that modeling is as much an art as it is a science. But like any good artist, you must master the techniques to use them correctly.

The next level of sophistication in regression models came with the advent of Generalized Linear Models (GLMs). GLMs allow for different types of response variable and some degree of non-linearity in the relationship between the response and explanatory variables. The relationship will still be linear at some level, but it might not be at the response level, it might only be linear at the level of the link function. What is the link function? It is a fundamental component of a GLM and is what allows responses to be constrained to a specific range of values. The link function, as its name implies, links the linear predictor and the response variable so that the model equation looks like:

$$g(E(Y)) = \alpha + X\beta, \quad (9.2)$$

where g is the link function, $E(Y)$ is the expected value of the response variable, and as in simple linear regression (see Eq. 9.1), α is the intercept (a constant), X is the predictor variable, and β is the regression coefficient. For a vector of n observations, the equation is in matrix form, where β is a vector of parameters and X is a matrix of predictor observations. The presence of a link function in Eq. (9.2) means that to obtain a prediction from this model, we need to apply the inverse of the link function to the linear predictors. As an example, consider a model with a log-link function. The inverse of the log is the exponent. This means that we need to exponentiate linear predictors to obtain the predicted value of Y for the corresponding values. But then, this also means that, irrespective of the covariate values and the coefficients estimated, the prediction will be positive (because the exponent of any number is positive). Some link functions allow values predicted for the response variable to be constrained (limited) to between 0 and 1, further increasing the range of modeling possibilities to include binary responses (e.g., presence/absence) or proportions. For instance, binary response variables like presence/absence are modeled using a binomial GLM, with logistic regression being a special case of a binomial GLM, where the link function is the logit function. Count data can be modeled using a Poisson GLM. The Poisson distribution is quite inflexible, however, because as noted above, it assumes that the mean and the variance are the same. Quite often, biological data are overdispersed, meaning that the variance is greater than the mean. For such count data, a quasi-Poisson or negative binomial response is often a second natural choice as it allows the variance to be greater than the mean. Finally, we could also consider other less commonly used, but equally useful, GLMs: (1) multinomial regression when the response can take one of several categorical outcomes, (2) gamma regression where the response is strictly positive, and (3) beta regression when the response is a probability or a proportion.

While GLMs allow added flexibility to standard linear regression as a result of the link function, if the relationship between the response and the predictors is highly non-linear (i.e., cannot be assumed linear even on the link function scale), then a GLM will not be adequate. This is where we need to bring non-linear functions into play, and perhaps the most widely used non-linear approach is the Generalized Additive Model (GAM). GAMs also consider a link function to allow different distributions for the response variable (as in GLMs), but we now have the response being a function of smooth functions of the predictors. In a univariate case, the model equation looks like:

$$g(E(Y)) = \alpha + f(x), \quad (9.3)$$

where g is the link function, $E(Y)$ is the expected value of the response variable, α is the intercept, x is the predictor variable, and f is a function such as a polynomial or spline. The polynomial or spline applies a smooth, curved-type function to the variable.

All the models described so far, be it a simple linear model (LM), a GLM, or a GAM, include only independent variables that are considered to be fixed effects. However, sometimes the inclusion of random effects might be necessary. A random effect is useful when we have observed a (random) subset of a larger population of possible values for a covariate. For example, a study may be interested in identifying responses of bats from a certain population before, during, and after exposure to high-frequency sound. The individual bats, whose responses were measured before, during, and after exposure, are a random effect. Random effects can be incorporated into a range of linear regression type models. For instance, Generalized Linear Mixed Models (GLMM) and Generalized Additive Mixed Models (GAMM) are GLMs and GAMs that incorporate both fixed and random effects. The reader is referred to Harrison et al. (2018) for an overview of mixed models in ecology, Pedersen et al. (2019) for non-linear models including mixed effects, and Nakagawa and Schielzeth (2010) for a review of the general issue of dealing with repeated

measurements sharing a correlation structure in biological studies.

Despite these advances, some data still do not fit the distributional requirements of GLMs and GAMs. Generalized Estimating Equations (GEEs) have been introduced recently, and hence they might still be considered in their infancy, but they are showing promising results. GEEs generalize GLMs and GAMs even further by not requiring that the response variable come from a particular family of distributions. GEEs simply impose a relationship between the mean and variance of the response. These models also allow a wide range of correlation structures to be imposed on the data, making them quite appealing when there are many observations clustered inside a few individuals. GEEs are marginal models in that the focus of inference is on the population average, and we are not so interested in the responses at the individual level. GEEs are quite specialized, and the reader is referred to Zuur et al. (2009, Chap. 12) for an introduction.

In addition to the somewhat “general” regression models above, there is a range of specialized regression models that are worth considering in certain biological questions. For instance, we have mentioned the problem of overdispersion. Often with biological data, we have very special cases of overdispersion in which there is an excess of zeroes. For example, consider you are trying to model the number of echolocation clicks a sperm whale produces per second as a function of depth, time of day, and sex. There are (at least) two reasons for there being zero clicks in a given second. A whale is in a silent state when recorded and many zeroes occur in successive seconds, or the whale is in a click-producing state but does not produce a click in the given second recorded. The regression models discussed above will likely fail to produce reasonable answers because the excess zeroes from the silent periods (potentially not explained by the covariates; i.e., not dependent on sex, depth, or time of day) cannot be accommodated. Under such a scenario, hurdle models or zero-inflated models might come in handy. While these are advanced methods and more difficult to implement and evaluate, they are worth knowing about. The reader is referred

to Martin et al. (2005) for a gentle introduction to the topic with ecological examples.

Truncated regression is another special case of regression under which some values of the response variable cannot be observed. An example is modeling animal group sizes as a function of their acoustic footprint (e.g., the number of sounds produced by a group that are detected per minute). Now that you know about GLMs, your first thought might be to consider a Poisson or negative binomial GLM, with group size as the response variable and numbers of sounds detected as the predictor. However, in modeling this, you soon face a problem: You fit your model and make some predictions, one of which is a group size of zero! What does this mean? Nothing really, it is what we call an inadmissible estimate and a clear sign that something is not adequate. Under such a case, you might want to try a zero-truncated regression, which is essentially a GLM for which zeroes cannot be observed. Chapter 11 in Zuur et al. (2009) explores both zero-inflated and zero-truncated models.

Survival models are regression techniques that deal with a special type of response variable: the time up to an event. While these types of models were developed to model survival of animals, plants, and people, they can be used in any scenario where observations might be censored. Censored data result when we do not know the real value of the response variable but know it is at least above or below some limit or within some interval; say because we observe an animal is dead at a given time, and/or we know it was alive at a different time. For example in a bioacoustic study, a researcher may wish to model the time animals take to produce their first acoustic cue, and animals are observed for 5 min each. However, we do not know when an animal produced a cue before observations began (i.e., left censoring). In addition, an animal might not produce any cues during the 5 min, or the animal might leave the study area before the 5 min elapse (i.e., right censoring). Finally, if we recorded only which minute, but not the actual second a sound was produced, we would only know that the event occurred sometime within the interval of that minute. These are interval

Table 9.5 Description of some commonly used models to test the association between multiple explanatory variables and a response variable

Model type	Use
Generalized Linear Modeling (GLM)	Allows different distributions for the response variable and some degree of non-linearity in the relationship between response and explanatory variables
Generalized Linear Mixed Effects Modeling (GLMM)	An extension of GLM for use with random effects (e.g., repeated measures of subjects)
Generalized Additive Modeling (GAM)	Allows different distributions for the response variable (as in GLMs) modeled as a function of smoothed predictors
Generalized Additive Mixed Effects Modelling (GAMM)	An extension of GAM for use with random effects (e.g., repeated measures of subjects)
Generalized Estimating Equations (GEE)	Do not require the response variable to come from a particular family of distributions, and allows correlation structures in the data to be accounted for

censored data. While a somewhat contrived example, this allows us to introduce the different kinds of censoring that are common in survival analysis.

Generalized Least Squares (GLS) is a regression approach that might be used when we want to relax the usual assumption of homogeneous residual variance by modeling the variance as a function of covariates. Zuur et al. (2009, Chap. 4) provide examples of the use of GLS and Reyier et al. (2014) give an acoustics application of GLS. Another perhaps more specialized use of such a regression technique is when we want to consider a general non-linear model with a specific form to relate a response variable with covariates. Then we might still want to find the parameters of the model that best fit the data. A way to do so is, akin to what might happen if one considers a straight line, to find the parameter values that minimize the sum of the squares of the residuals (i.e., the difference between the observations and the model). In a simple regression context, the model produces the fitted line, while in a generalized least squares context, the model is any function in which we might be interested. For example, if you want to determine the propagation loss (PL) for a sound that has traveled from the source to the receiver, and you expect it is proportional to $\log(r)$, where r is the range, then your model is $PL = K \log(r)$. Based on measurements of received levels of sounds with known source level, you may apply a GLS regression to estimate the value of K that best fits your data. If K is close to 10, then your environment supports cylindrical spreading, if it is close to

20, then sound is predicted to spread spherically (see Chaps. 5 and 6 on sound propagation in air and under water, respectively).

All the models described so far do not consider predictor variables that are in hierarchies. Hierarchical data occur when variables are nested within each other (i.e., organized into levels). For example, individuals from different resident populations can be said to be nested within subpopulations. In turn, subpopulations can be nested within populations. Hierarchical modeling (also known as multilevel modeling) is used when inferences need to be drawn for population means at specified levels and is useful for fitting models to data obtained from complex, multilevel survey designs. For example, a study may evaluate vocal complexity of elephants at the population, sub-population, and resident population levels. Here, we do not discuss these methods further. Rather, we refer the reader to Cressie et al. (2009) and Royle and Dorazio (2008) for descriptions of these methods, including their strengths and limitations.

Given the large range of models available (a taste of which has been described above), what should aspiring ecologists today have in their statistical regression toolbox? We propose that a bare minimum is an understanding of the structure, implementation, outputs, and interpretation of GLMs, GLMMs, GAMs, and GAMMs (Table 9.5). Parameter estimates and significance tests resulting in p -values are common outputs of software capable of fitting GLMs, GLMMs, GAMs, GAMMs, and GEEs. For a practical guide to applying these in behavioral and

ecological studies, see Zuur et al. (2009). O'Hara (2009) and Bolker et al. (2009) provide good introductions to GLMMs for ecologists, and the books by Zuur et al. (2007, 2009) provide information to implement and interpret GLMMs. For GAMs, the book by Wood (2006) is a standard reference, and Zuur et al. (2009) has worked-out examples in the software R.

Most of the models described in this section can be implemented in a frequentist framework, for instance using maximum likelihood or restricted maximum likelihood estimation. Nonetheless, for more complex models such as those including (often complex) spatial and temporal covariates (i.e., spatio-temporal models), Bayesian implementations are gaining ground. For instance, GLMs and GLMMs are fitted via maximum likelihood, or Markov Chain Monte Carlo (MCMC). MCMCs are Bayesian iterative solutions and are described in Gamerman (1997), Brémaud (1999), Draper (2000), and Link (2002). With advances of widely available implementations, users might even be using Bayesian approaches without realizing it. An example is the Integrated Nested Laplace Approximation (INLA) implemented via R-INLA (www.r-inla.org) and its derivatives that allow fitting complex spatio-temporal models without the Bayesian framework being obvious (by not requiring priors to be explicitly defined). The philosophical nuances of which framework might be more adequate under given settings, however, are beyond what we hope to discuss in this chapter.

9.5.3.1 Model Validation, Selection, and Averaging

Depending upon whether modeling is undertaken for explanatory or predictive purposes, approaches for model validation and selection may differ (Shmueli 2010). Validation means that the model has been demonstrated to have satisfactory accuracy for its intended use (Rykiel Jr 1996). Validation in explanatory modeling commonly takes the form of goodness-of-fit and residual diagnostics. Goodness-of-fit tests evaluate how well-observed values agree with those expected under the statistical model (Maydeu-

Olivares and Garcia-Forero 2010), while residual diagnostics determine whether residuals fit the assumption of being effectively random (see Zuur et al. 2009 for common examples in ecology). Checking for multi-collinearity (i.e., collinearity between two or more covariates) is also standard for explanatory modeling, while it is close to irrelevant for predictive modeling (see Shmueli 2010 for detailed discussion). In contrast to explanatory modeling, model validation in predictive modeling is focused on evaluating the model's ability to generalize and predict new data. Validation commonly is undertaken using approaches such as cross-validation. In cross-validation, the model's ability to accurately predict a new data set is assessed after calibrating it with a training dataset (Shmueli 2010; Cawley and Talbot 2010).

Once a set of models have been validated, the best candidate model is selected (though model validation and selection can often be an iterative process). Approaches to model selection, again, depend upon whether modeling has an explanatory or predictive goal. In explanatory modeling, the explanatory power of nested candidate models is commonly compared with a step-wise approach using significance testing (e.g., using an F -test). Here a nested model refers to one composed of subsets of covariates of another candidate model. Caution should be taken, however, as researchers may be inclined to remove covariates that are not significant, even when there is a strong theoretical justification for retaining them since they are relevant in the models, regardless of whether they are significant or not (Shmueli 2010). For example, a covariate representing the age class of a sparrow in a study assessing the influence of predator presence on sparrow vocal behavior may be of theoretical importance in the model. Model selection in predictive modeling commonly involves a priori specification of candidate models and selecting the best model based on the smallest possible number of parameters that adequately represent the data (i.e., the principle of parsimony). The simpler a model is, the more it can be generalized, while more complex models (containing more parameters) are more specific to the data used to fit the model. Consequently,

criteria for model selection have been developed that essentially maximize the likelihood while penalizing for the number of parameters included. The Akaike's Information Criterion (AIC; see Akaike 1974) and Bayesian Information Criterion (BIC) currently are the most commonly used, among a range of others available. They are widely used for comparing nested and non-nested models (Burnham and Anderson 2002), although there is some discussion around suitability for use in non-nested models (see Ripley 2004). Resulting criteria such as AIC or BIC values for candidate models are then compared and the model yielding the lowest value is generally deemed to be preferred. Note that there is active research on the circumstances under which AIC, BIC, and the many other criteria available perform best, and whether they should be used together to inform model selection (Kuha 2004). An important take-home message is that model selection criteria such as AIC and BIC can only suggest a preferred model from those compared, even if they all perform poorly at the validation stage. In other words, the preferred model may still be a poorly fitting model, and therefore, selection criteria are only relative measures of model goodness-of-fit.

In predictive modeling, averaging over a range of plausible models has become widely used to reduce prediction error and improve model selection uncertainty. This is undertaken, for example, by computing a measure that ranks the set of plausible models according to their support by the data (e.g., Akaike weights), applying the weights to predictions from each model, and then computing the average. This provides weighted averaged predictions, with weights dependent on how much each model is supported by the data. There are many other methods for undertaking model averaging. Model averaging performance depends on each model's predictive bias and variance and covariance between models, among other things (see McElroy 2016 for complete discussion). In recent work, model averaging has been shown to be particularly useful when predictive errors of contributing model predictions are dominated by variance, and when

covariance between models is low (McElroy 2016).

While a highly simplified overview of some tools available on the topic of model validation, selection, and averaging has been provided here, researchers should be familiar with them and access the latest literature to identify the appropriate approaches for their study.

9.5.4 The Future of Bioacoustical Analytical Approaches

In this chapter, we have only provided a flavor of common approaches used today and have not delved into the wide range of new developments being introduced into the discipline. Interdisciplinary research linking the fields of biology, ecology, and statistics has a long tradition of providing fertile ground for innovative statistical methods, with many methods having been developed when existing methods were not adequate to cope with new problems (Olivier et al. 2014). The current revolution in data acquisition systems (see Chap. 2), such as high-resolution sensors in animal-borne tags and increasing numbers of long-term passive acoustic deployments that lead to big data, is also likely to influence the next generation of statistical methods suited for ecological and acoustical analysis. Analysis of big data through increased computational capacity has already provided a range of new powerful tools to science.

As an example of such approaches, machine learning is rapidly gaining in popularity as it increasingly improves pattern recognition accuracy (Christin et al. 2019). Such methods can improve processing capacity in large datasets resulting from acoustic instrumentation. An example of more sophisticated analytical approaches is the growing use of hierarchical, state-space, and hidden process methods (e.g., Auger-Méthé et al. 2020 for an introduction to their application in ecology) that model underlying processes while accounting for biases and uncertainty. Advances in these approaches may improve our ability to predict future scenarios and implement intervention before a potentially

undesirable future scenario unfolds (see Cressie et al. 2009 for discussion).

We also suggest readers to be acquainted with the growing work being conducted in the area of statistical decision theory, which is concerned with making decisions by accounting for uncertainties involved in the decision process using statistical knowledge resulting from data collected. Rather than attempting to provide a general review of the large field of decision theory here, we refer the reader to an introduction in its application to ecology by Williams and Hooten (2016), which will introduce the reader to a range of other resources on the topic.

Because the advancement of these and many other methods are continually evolving, researchers are encouraged to keep well-informed of current developments appearing in methods-based scientific journals, such as *Methods in Ecology and Evolution*.

9.6 Examples in Bioacoustics

The wide range of quantitative approaches introduced above can be used to analyze bioacoustical data to answer research questions ranging from understanding natural vocal behavior to activity patterns, community and conservation ecology, habitat use, species diversity, distribution, occupancy, density and abundance, and anthropogenic impacts (among many others). Faunal groups that have been the subject of bioacoustics research include invertebrates, anurans (i.e., frogs and toads), fish, birds, bats, other terrestrial mammals, and marine mammals, but many others could be considered. As long as sound is produced, it could be used as a source of information. A recent review documented 460 peer-reviewed published papers on passive acoustic monitoring in terrestrial habitats alone, with bats (50% of papers) and activity patterns (24%) dominating (Moreria Sugai et al. 2018). Marine mammals feature prominently in bioacoustic research as water is a highly conducive medium for sound to travel through, and visual observations can prove comparatively expensive for limited returns on detections.

Rather than reviewing analytical approaches across the hundreds of existing bioacoustics studies, we have selected two recent studies as examples, and discuss the rationale for the particular analytical approaches taken. The research topics in the example studies are exploring temporal changes in call frequency and using acoustic data for abundance and density estimation.

9.6.1 Temporal Changes in Call Frequency

As indicated previously, due to ever-increasing computing power and storage and technological advances in acoustic equipment, acoustic studies can provide extremely long-term datasets. These datasets allow us to explore changes to calling behavior on a scale that, until recently, would have been very difficult. A recent example is illustrated in Miksis-Olds et al. (2018) where the frequency content of a type of blue whale song recorded primarily in the Indian Ocean was investigated. The song type is attributed to a pygmy blue whale subspecies (*Balaenoptera musculus indica*, Committee on Taxonomy 2021) that appears to be resident in the northern Indian Ocean. The song type has three distinct units, and this analysis focused on the ~60-Hz component of Unit 2, a frequency-modulated upsweep, and Unit 3, a ~100-Hz tonal downsweep. A decade of data from the Indian Ocean Comprehensive Nuclear-Test-Ban Treaty International Monitoring Station (CTBTO IMS) at Diego Garcia was analyzed (2002–2013). Ambient noise was also analyzed, but we do not focus on that part of the study here.

Power spectral densities (PSD) were computed for 2-h sections of data, which could be used to detect peaks in the frequency bands of interest (approximately 56–63 Hz for the 60-Hz component of Unit 2, and 107–100 Hz for Unit 3), using a 3-dB signal-to-noise threshold. The paper shows a figure of number of hours with vocal presence detected each week, for each year (Fig. 9.3 in Miksis-Olds et al. 2018), highlighting the importance of producing exploratory plots; in this case, the variability in the data is made clear.

The average over each week, across years, was used to identify weeks with peak average vocal presence. Weeks 21 and 22 were those with peak average vocal presence and data from these weeks were investigated further. The frequency peaks from the PSDs from these weeks across all years were measured. A linear regression model was fitted to the week 21 and 22 frequency peak measurements from all years. The response variable was frequency, and year and song unit were explanatory variables. Song unit was included in the model as a factor variable. An interaction was also included between year and song unit, which was used to investigate whether the rate of any frequency change over time differed between the two song units. Model assumptions (linearity, constant error variance, error independence, and normality) were all assessed using diagnostic plots and relevant hypothesis tests, and all model assumptions were met.

The linear model results are depicted in Fig. 9.10. The figure shows all weekly data plotted (blue dots) with the modeled 21–22 week data highlighted in red for both song units. Again, the utility of plotting data is clear here: the decline in frequency is evident, with an apparent difference in rate of decline between the two units. The linear model results confirmed the frequency decline; the frequency of the ~60-Hz Unit 2 decreased at a rate of 0.18 Hz/year, while the frequency of Unit 3 decreased at 0.54 Hz/year. The interaction term was selected during model selection (using an F-test), which confirmed that the rates of frequency decline were indeed different between the two units.

This analysis shows that simple regression analyses can be very effective in confirming patterns observed in exploratory data plots. We note here that the regression analysis in the paper focused on data from weeks 21 and 22 to be comparable with methods from a similar study (Gavrilov et al. 2012). However, frequency measurements were taken across all weeks of each year (as shown in Fig. 9.10), which could also be used in a regression model. In addition, it is common for bioacoustical analyses to have several natural extensions. In this case, relaxing the Gaussian assumption could be considered via

a Generalized Linear Model, or non-linear patterns in the frequency decline could be explored using a Generalized Additive Model.

9.6.2 Abundance and Density Estimation

The estimation of animal population size (abundance) and the number of animals in a given area (density) are metrics that are very informative for management and conservation actions. There are several abundance and density estimation methods available (e.g., Borchers et al. 2002); popular methods include mark-recapture and distance sampling. Such methods are known as absolute abundance or density estimation methods, as the methods estimate the total number of animals (in a defined area, for density estimates), including animals missed by a survey. Common reasons why animals are not detected during a survey is that they may be too far away, and/or detection is made difficult by environmental conditions (e.g., rough seas may prevent marine mammal sightings at sea unless the animals are very close, or windy conditions may mask the sounds of singing birds in recordings). The probability of detecting an animal is a key parameter in absolute abundance and density estimation methods, and accounts (in part) for undetected animals during a survey.

Acoustic data are increasingly being used for absolute abundance and density estimation, both in terrestrial and marine environments (e.g., Marques et al. 2013; Stevenson et al. 2015). Here we discuss a density estimation analysis for Blainville's beaked whales (*Mesoplodon densirostris*) from seafloor-moored hydrophone data recorded in the Bahamas (Marques et al. 2009). The analysis involved several of the concepts we have discussed throughout the chapter, which we highlight here.

The paper begins by introducing the density estimation equation (i.e., the estimator; see Sect. 9.4.2). The equation contains several parameters to be estimated, including the probability of detecting a beaked whale echolocation click on one of the seafloor-moored hydrophones. Survey

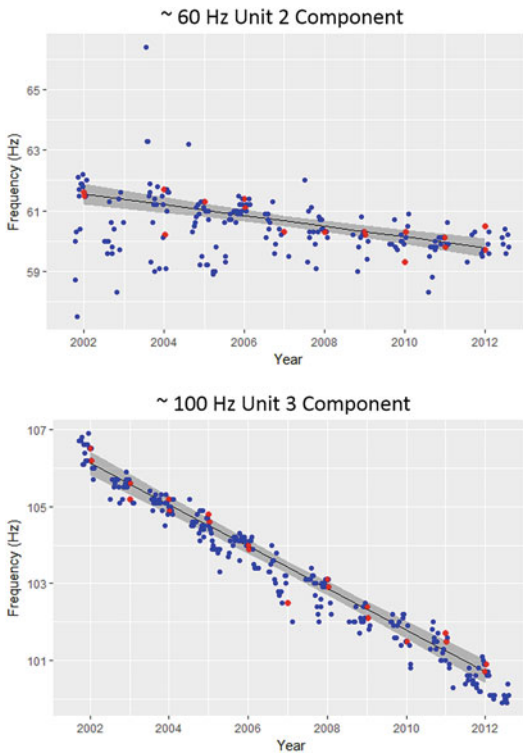


Fig. 9.10 Peak frequency of Sri Lankan whale vocalizations determined from weekly PSD sound averages. The blue circles are the weekly peaks measured throughout the season when whales were vocally present. The trend line is related to the red circles that are peak frequency from weeks 21 and 22 of each year. The greyed regions designate the 95% confidence intervals for the trend. Reprinted with permission from Miksis-Olds et al. (2018). © Acoustical Society of America, 2018. All rights reserved

design and variance estimation of the parameters (including confidence intervals) are also discussed. A summary of methods to estimate the detection probability is given. Mark-recapture and distance sampling methods are commonly used approaches to estimate the detection probability, but Marques et al. (2009) needed an alternative method, given that the hydrophone recordings were not suitable for either mark-recapture, or distance sampling-based methods. Therefore, a trial-based detection probability estimation method was used. The specific trial-based method used in this study relied on auxiliary data from animals tagged with acoustic tags, which

swam near the moored hydrophones. Clicks produced by the animals and recorded on the tags created “trials”; a successful trial was achieved if the same clicks recorded on tags of the tagged animal were detected on the moored hydrophones. In addition, the tag data provided the slant distance of each tagged animal from the moored hydrophones, as well as the animal’s orientation toward, or away from, a given moored hydrophone. These data allowed detection probability to be modeled as a function of a whale’s orientation and distance from the moored hydrophones using regression modeling. Specifically, a Generalized Additive Model (GAM) was used due to its flexibility in allowing non-linear relationships between the response and explanatory variables. The response variable was defined as the detection, or non-detection, of each click produced by the tagged animal on the moored hydrophones. The explanatory variables, or covariates, were (a) the horizontal off-axis angle (hoa) and (b) vertical off-axis angle (voa) of the tagged whale, with respect to a given moored hydrophone, and (c) the distance of the tagged whale from the hydrophone. A binomial distribution was assumed for the response variable due to the binary nature of the trial data (i.e., detected, or not detected) and a logistic link function was used in the GAM. Finally, to estimate the average detection probability (i.e., a single parameter value for the estimator), a Monte Carlo simulation was implemented where the dive profiles from the tags were randomly placed around virtual moored hydrophones. In the simulation, the slant range and orientation of the clicks from the dive profiles from the moored hydrophones could be calculated, and then these values could be used along with the GAM to predict the detection probability for each click in the simulation. The average of these predicted detection probabilities was used in the estimator. Two other parameters required for the estimator, the false-positive proportion and cue production rate, are discussed in the paper in detail, on which we do not focus here.

The results of the GAM are shown in Fig. 9.11. The modeled relationships between (a) detection probability and slant range, (b) vertical and horizontal off-axis angle and

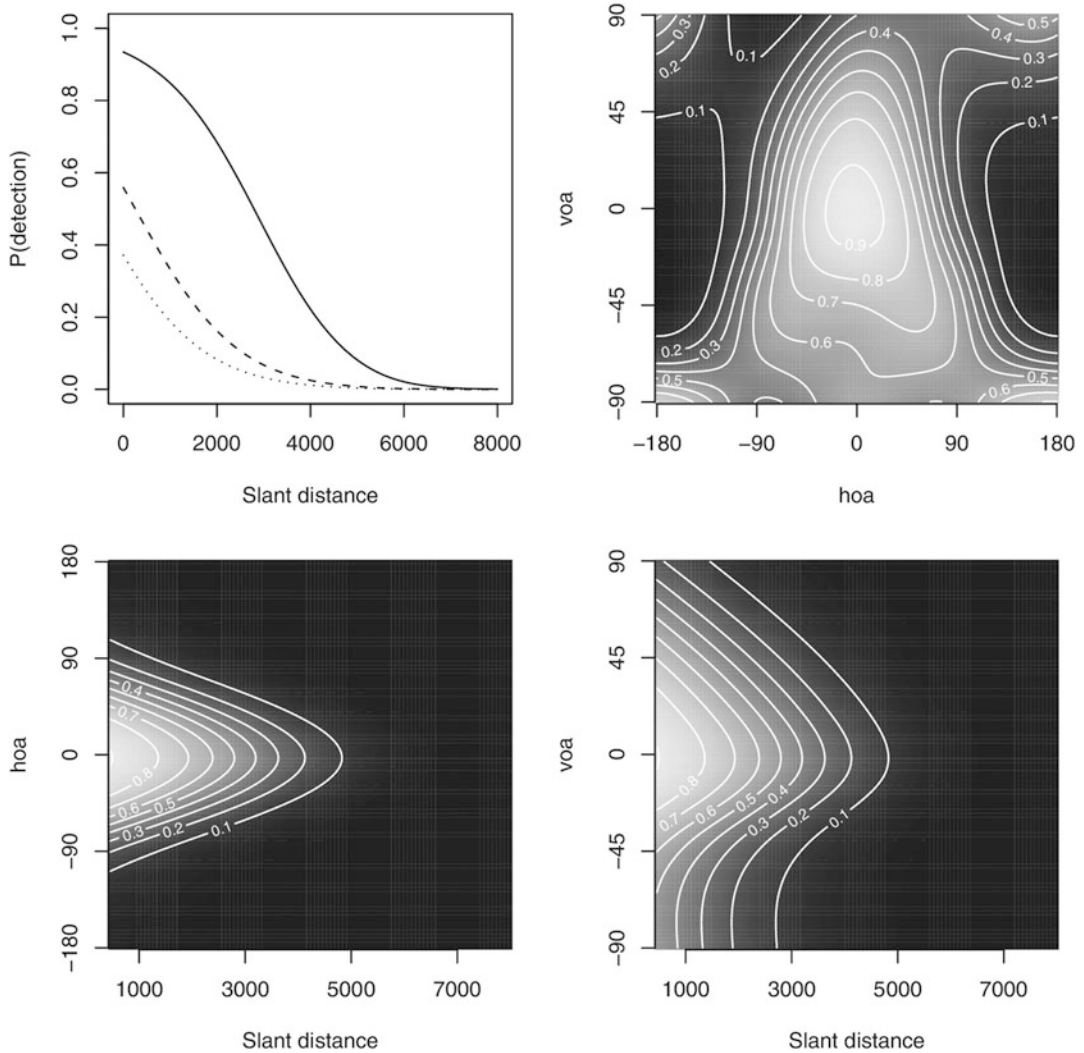


Fig. 9.11 The estimated detection function. Plots (on the response scale) of the fitted smooths for a binomial GAM model with slant distance and a 2D smooth of *hoa* and *voa*. For the top left plot, the off-axis angles are fixed at 0, 45, and 90° (respectively the solid, dashed, and dotted lines). Remaining plots are two-dimensional representations of

the smooths, where black and white represent respectively an estimated probability of detection of 0 and 1. Distance (top right panel) and angle not shown (bottom panels) are fixed respectively at 0 m and 0°. Reprinted with permission from Marques et al. (2009). © Acoustic Society of America, 2009. All rights reserved

detection probability, (c) horizontal off-axis angle and slant range, and (d) vertical off-axis angle and slant range are all depicted. The average detection probability of a beaked whale click within 8 km of a moored hydrophone was estimated to be 0.03 (i.e., if a beaked whale click was produced within 8 km of a moored hydrophone, the study estimated that there was, on average, a 3% chance

of detecting that same click). The variance around the average was estimated using the bootstrap and presented as a coefficient of variation (CV, defined in Sect. 9.4.2) and was estimated to be 0.16, or 16% when expressed as a percentage. Finally, the estimator was used to estimate beaked whale density in the study area of either 25.3 (CV: 19.5%) or 22.5 (19.6%) animals per 1000 km²,

depending on the false-positive proportion used (two estimates were produced using differing methods).

9.7 Software for Analyses

There are many standard, relatively easy-to-use software packages that require no (or very little) coding skills to carry out statistical analyses, including *SPSS* (IBM Corp., Armonk, NY, USA), *Statistica* (TIBCO Software, CA, USA), *Stata* (StataCorp, College Station, TX, USA), *Minitab* (Minitab Inc., State College, PA, USA), *Xlstat* (Addinsoft, Ile-de-France, France), and *SAS* (SAS Institute, Cary, NC, USA), among others. In the field of bioacoustics, it is common for acoustic data to be processed in *MATLAB* (The MathWorks Inc., Natick, MA, USA) due to its powerful signal processing package. *MATLAB* users may find that their workflow is streamlined by undertaking statistical analyses in the same software if all required tools are available.

For those planning, however, on undertaking analyses that draw from the most recent up-to-date developments in statistical ecology and require a highly flexible environment to do so, a free open-source software environment like R is recommended (R Core Team 2020). R is primarily used for statistical computing and production of graphics (though R's GIS, and even signal processing capabilities, are expanding). The software benefits from a large number of base and contributed packages that can easily be downloaded and an environment in which users may develop their own algorithms and packages. There are now many sources of instructional manuals and books guiding users on how to create high-quality data representations and run analyses in R, including Crawley (2013), Kerns (2010), Zuur et al. (2009), Bolker (2008), Lawson (2014), among many others. The *CRAN Task View: Analysis of Ecological and Environmental Data*¹ maintained by Gavin Simpson is an excellent resource for locating suitable packages

¹ CRAN Task View: <https://CRAN.R-project.org/view=Environmetrics>; accessed 9 November 2020.

for statistical analysis of biological data. R can be accessed and downloaded through a web browser² and for most users, we recommend a user-friendly GUI like *RStudio* (RStudio Team 2020³). *RStudio* is an integrated development environment for R that includes a console, an editor for code development and execution, and tools for plotting, debugging, tracking history, and managing the workspace. An interesting feature of R integrated with *RStudio* is the ability to adhere in a straightforward way to the concept of reproducible research via dynamic reports in *RMarkdown*. If the reader is new to the topic, we recommend the book by Xie et al. (2020).⁴

9.8 Summary

A key outcome of bioacoustics research is the production of new knowledge that informs conservation management. The knowledge produced needs to be reliable and easily understood, which is no trivial task given the complicated nature of animal behavior. The reality is that the phenomena from which we want to derive inferences are multifaceted, with many interconnecting attributes, and patterns and signals obscured by statistical noise (i.e., variability not associated with the conditions under investigation). Consequently, underlying mechanisms that explain the patterns we observe are not easily revealed.

Not only are animal behaviors occurring in a highly complex environment, but many challenges are presented in conducting the research itself. For instance, as researchers we are not easily able to avoid or reduce the statistical noise in the environment by controlling field conditions; and when we undertake experiments of animals in captivity to reduce noise in a laboratory, we cannot be sure that results are

² R Core Team is accessible at <https://www.r-project.org/>; accessed 1 January 2020.

³ *RStudio* is accessible at <https://www.rstudio.com/products/RStudio/>; accessed 9 November 2020.

⁴ *RMarkdown: The Definitive Guide* by Xie Y, Allaire JJ, Golemund G: <https://bookdown.org/yihui/rmarkdown/>; accessed 9 November 2020.

transferable to the wild. In addition, we introduce biases in our observations through our own subjective, non-random filters. Only by understanding these filters can we either eliminate or adjust biases to make reliable inferences about nature.

Quantitative skills, including survey design considerations, are therefore an essential part of a bioacoustician's toolkit and should be viewed just as essential as field skills and signal processing methods. These statistical methods are tools that enable the researcher to ask difficult but often important and exciting questions about their research topic.

However, given the complexity in nature, research design challenges, and the multi-disciplinary nature of studying animal behavior through acoustics, it is not realistic to expect specialists in one field to become experts across multiple fields (i.e., behavior, ecology, bioacoustics, and statistics). What behaviorists and bioacousticians can aim for is to understand foundational statistical concepts, have a broad knowledge of the range of existing techniques available, and be able to identify critical pitfalls in survey design and data analyses. In addition, practitioners should be able to conduct a range of current standard analyses and know when to seek support for more sophisticated approaches.

It is our hope that through the introduction of basic statistical concepts in this chapter, readers can more confidently avoid design and analysis pitfalls and make the necessary considerations to select the most suitable approaches to successfully answer their research questions. We would like researchers to feel empowered to critically evaluate the transferability of standard practices across broader spectra of questions and identify inadequacies where they occur. Finally, and foremost, we hope that at the conclusion of this chapter, readers feel inspired to place greater focus on the biological significance of research outputs, using quantitative methods as a tool to support their conclusions.

We close this chapter by providing you, the reader, with our culinary rendition of the meaning of statistics: It is the science that uses data as its main ingredient, uncertainty as a key seasoning

driving the final flavor of a meal, and guides the collection and mixing of the ingredients, through sampling, experimentation, and analysis. Taken together, hopefully, delicious scientific meals will result, by drawing meaningful and reliable inferences from data. Statistics is paramount for science in general, and bioacoustics is in that regard no exception.

Acknowledgement We thank Steve Buckland and Jay Barlow for their helpful comments prior to Springer's peer-review.

References

- Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Autom Control* 19(6): 716–723
- Auger-Méthé M, Newman K, Cole D, Empacher F, Gryba R, King AA, Leos-Barajas V, Flemming JM, Nielson A, Petris G (2020) An introduction to state-space modeling of ecological time series. arXiv preprint arXiv:2002.02001
- Beninger PG, Boldina I, Katsanevakis S (2012) Strengthening statistical usage in marine ecology. *J Exp Mar Biol Ecol* 426:97–108
- Bolker BM (2008) Ecological models and data in R. Princeton University Press, Princeton
- Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH, White JSS (2009) Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol Evol* 24(3):127–135
- Borcard D, Gillet F, Legendre P (2011) Numerical ecology with R. Springer, New York
- Borchers DL, Buckland ST, Zucchini W (2002) Estimating animal abundance. Springer, New York
- Box GE (1976) Science and statistics. *J Am Stat Assoc* 71(356):791–799
- Brémaud P (1999) Markov chains: Gibbs fields and Monte Carlo simulation. Springer, New York, pp 253–322
- Brown A, Smith J, Salgado Kent C, Marley S, Allen S, Thiele BL, Erbe C, Chabanne D (2017). Relative abundance, population genetic structure and passive acoustic monitoring of Australian snubfin and humpback dolphins in regions within the Kimberley. <https://doi.org/10.13140/RG.2.2.17354.06082>
- Burnham K, Anderson D (2002) Model selection and multimodel inference: a practical information-theoretic approach, 2nd edn. Springer, New York
- Casella G, Berger RL (2002) Statistical inference. Duxbury, Belmont, CA
- Cawley GC, Talbot NL (2010) On over-fitting in model selection and subsequent selection bias in performance evaluation. *J Mach Learn Res* 11:2079–2107

- Christin S, Hervet É, Lecomte N (2019) Applications for deep learning in ecology. *Methods Ecol Evol* 10(10): 1632–1644
- Cochran WG (1977) *Sampling techniques*, 3rd edn. Wiley, New York
- Cohen J (1988) *Statistical power analysis for the behavioral sciences*, 2nd edn. L. Erlbaum Associates, Hillsdale, NJ
- Cohen J (2013) *Statistical power analysis for the behavioral sciences*, 2nd edn. Routledge, New York
- Committee on Taxonomy (2021) List of marine mammal species and subspecies. Society for Marine Mammalogy. www.marinemammalscience.org. Accessed 2 Sep 2021
- Crawley MJ (2013) *The R book*, 2nd edn. Wiley, Hoboken, NJ
- Cressie N, Calder CA, Clark JS, Hoef JMV, Wikle CK (2009) Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. *Ecol Appl* 19(3):553–570
- Dytham C (2011) *Choosing and using statistics: a biologist's guide*, 3rd edn. Wiley, Boca Raton, FL
- Ellis PD (2010) The essential guide to effect sizes: statistical power, meta-analysis, and the interpretation of research results. Cambridge University Press, Cambridge
- Ellison AM (2004) Bayesian inference in ecology. *Ecol Lett* 7(6):509–520
- Ellison AM, Gotelli NJ, Inouye BD, Strong DR (2014) P values, hypothesis testing, and model selection: it's déjà vu all over again. *Ecology* 95(3):609–610
- Everitt B, Hothorn T (2011) *An introduction to applied multivariate analysis with R*. Springer, New York
- Fisher RA (1959) *Statistical methods and scientific inference*, 2nd ed. Oliver and Boyd, Edinburgh, UK
- Ford ED (2000) *Scientific method for ecological research*. Cambridge University Press, Cambridge
- Gamerman D (1997) Sampling from the posterior distribution in generalized linear mixed models. *Stat Comput* 7(1):57–68
- Gavrilov AN, McCauley RD, Gedamke J (2012) Steady inter and intra-annual decreases in the vocalization frequency of Antarctic blue whales. *J Acoust Soc Am* 131:4476–4480. <https://doi.org/10.1121/1.4707425>
- Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2013) *Bayesian data analysis*, 3rd edn. CRC Press, Boca Raton, FL
- Harrison XA, Donaldson L, Correa-Cano ME, Evans J, Fisher DN, Goodwin CE, Robinson BS, Hodgson DJ, Inger R (2018) A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ* 6:e4794
- Hilborn R, Mangel M (1997) *The ecological detective: confronting models with data*, vol 28. Princeton University Press, Princeton
- Kerckhoff AJ, Enquist BJ (2009) Multiplicative by nature: why logarithmic transformation is necessary in allometry. *J Theor Biol* 257(3):519–521
- Kerns GJ (2010) *Introduction to probability and statistics using R*, 1st edn. G. Jay Kerns, Youngstown
- Kuha J (2004) AIC and BIC: comparisons of assumptions and performance. *Sociol Methods Res* 33(2):188–229
- Lawson J (2014) *Design and analysis of experiments with R*, vol 115. CRC Press, Boca Raton, FL
- Leek JT, Peng RD (2015) What is the question? *Science* 347(6228):1314–1315
- Link RF (2002) Principal applications of Bayesian methods in actuarial science: a perspective. *North Am Actuarial J* 6(2):129
- Manly BFJ (2007) *Randomization, bootstrap and Monte Carlo methods in biology*. CRC Press, Boca Raton, FL
- Manly BF, Alberto JAN (2014) *Introduction to ecological sampling*. CRC Press, Boca Raton, FL
- Marques TA, Thomas L, Ward J, Dimarzio N, Tyack PL (2009) Estimating cetacean population density using fixed passive acoustic sensors: an example with Blainville's beaked whales. *J Acoust Soc Am* 125(4): 1982–1994. <https://doi.org/10.1121/1.3089590>
- Marques TA, Thomas L, Martin SW, Mellinger DK, Ward JA, Moretti DJ, Harris D, Tyack PL (2013) Estimating animal population density using passive acoustics. *Biol Rev* 88(2):287–309
- Martin TG, Wintle BA, Rhodes JR, Kuhnert PM, Field SA, Low-Choy SJ, Tyre AJ, Possingham HP (2005) Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. *Ecol Lett* 8(11):1235–1246
- Matthiopoulos J (2010) *How to be a quantitative ecologist*. Wiley, Hoboken, NJ
- Maydeu-Olivares A, Garcia-Forero C (2010) Goodness-of-fit testing. *Int Encycl Educ* 7(1):190–196
- McCarthy MA (2007) *Bayesian methods for ecology*. Cambridge University Press, Cambridge
- McElreath R (2020) *Statistical rethinking: a Bayesian course with examples in R and Stan*. CRC Press, Boca Raton, FL
- McElroy TS (2016) Nonnested model comparisons for time series. *Biometrika* 103(4):905–914
- McGarigal K, Cushman SA, Stafford S (2000) *Multivariate statistics for wildlife and ecology research*, 1st edn. Springer, New York
- Miksis-Olds JL, Nieuwkirk SL, Harris DV (2018) Two unit analysis of Sri Lankan pygmy blue whale song over a decade. *J Acoust Soc Am* 144(6):3618–3626
- Moreria Sugai L, Freire Silva T, Wagner Ribeiro J, Llusia D (2018) Terrestrial passive acoustic monitoring: review and perspectives. *Bioscience* 69. <https://doi.org/10.1093/biosci/biy147>
- Nakagawa S, Schielzeth H (2010) Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. *Biol Rev* 85:935–956
- Nuzzo R (2014) Scientific method: statistical errors. *Nat News* 506(7487):150
- O'Hara RB (2009) How to make models add up—a primer on GLMMs. *Annales Zoologici Fennici*, *BioOne*, pp 124–137
- O'Hara R, Kotze J (2010) Do not log-transform count data. *Nat Precedings* 1:118–122
- Ortega A, Navarrete G (2017) Bayesian hypothesis testing: an alternative to null hypothesis significance

- testing (NHST). In: Psychology and social sciences. IntechOpen
- Paiva EG, Salgado Kent CP, Gagnon MM, McCauley R, Finn H (2015) Reduced detection of Indo-Pacific bottlenose dolphins (*Tursiops aduncus*) in an inner harbour channel during pile driving activities. *Aquat Mamm* 41(4):455–468
- Panzeri S, Magri C, Carraro L (2008) Sampling bias. *Scholarpedia* 3(9):4258
- Pedersen EJ, Miller DL, Simpson GL, Ross N (2019) Hierarchical generalized additive models in ecology: an introduction with mgcv. *PeerJ* 7:e6876
- Quinn GP, Keough MJ (2002) Experimental design and data analysis for biologists. Cambridge University Press, Cambridge
- Rahlf T (2019) Data visualisation with R: 111 examples. Springer Nature, New York
- Reyier EA, Franks BR, Chapman DD, Scheidt DM, Stolen ED, Gruber SH (2014) Regional-scale migrations and habitat use of juvenile lemon sharks (*Negaprion brevirostris*) in the US South Atlantic. *PLoS One* 9(2):e88470
- Ripley BD (2004) Selecting amongst large classes of models. In: Methods and models in statistics, In Honour of Professor John Nelder, FRS. World Scientific, New York, pp 155–170
- Royle JA, Dorazio RM (2008) Hierarchical modeling and inference in ecology: the analysis of data from populations, metapopulations and communities. Elsevier, New York
- Rykiel EJ Jr (1996) Testing ecological models: the meaning of validation. *Ecol Model* 90:229–244
- Salkind NJ (2010) Encyclopedia of research design, vol 1. Sage, Thousand Oaks, CA
- Shmueli G (2010) To explain or to predict? *Stat Sci* 25(3): 289–310
- Stauffer HB (2007) Contemporary Bayesian and frequentist statistical research methods for natural resource scientists. Wiley, Hoboken, NJ
- Stevenson BC, Borchers DL, Altwegg R, Swift RJ, Gillespie DM, Measey GJ (2015) A general framework for animal density estimation from acoustic detections across a fixed microphone array. *Methods Ecol Evol* 6(1):38–48. <https://doi.org/10.1111/2041-210X.12291>
- RStudio Team (2020) RStudio: integrated development for R. RStudio, PBC, Boston, MA. <http://www.rstudio.com/>
- R Core Team (2020) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. <https://www.R-project.org/>
- Touchon JC, McCoy MW (2016) The mismatch between current statistical practice and doctoral training in ecology. *Ecosphere* 7(8):e01394
- Underwood AJ (1997) Experiments in ecology: their logical design and interpretation using analysis of variance. Cambridge University Press, Cambridge
- Van Der Maaten L, Postma E, Van den Herik J (2007) Dimensionality reduction: a comparative review. *J Mach Learn Res* 10(66–71):13
- Warren VE, Marques TA, Harris D, Thomas L, Tyack PL, Aguilar de Soto N, Hickmott LS, Johnson MP (2017) Spatio-temporal variation in click production rates of beaked whales: implications for passive acoustic density estimation. *J Acoust Soc Am* 141(3): 1962–1974
- Wasserstein RL, Schirm AL, Lazar NA (2019) Moving to a world beyond “ $p < 0.05$ ”. Taylor & Francis, New York
- Wilcox RR (2010) Fundamentals of modern statistical methods: substantially improving power and accuracy. Springer, New York
- Williams PJ, Hooten MB (2016) Combining statistical inference and decisions in ecology. *Ecol Appl* 26: 1930–1942
- Xie Y, Allaire JJ, Golemund G (2020) R markdown: the definitive guide. CRC Press, Boca Raton, FL
- Yoccoz NG (1991) Use, overuse, and misuse of significance tests in evolutionary biology and ecology. *Bull Ecol Soc Am* 72(2):106–111
- Zimmer WM, Johnson MP, Madsen PT, Tyack PL (2005) Echolocation clicks of free-ranging Cuvier’s beaked whales (*Ziphius cavirostris*). *J Acoust Soc Am* 117(6):3919–3927
- Zuur A, Ieno EN, Smith GM (2007) Analyzing ecological data. Springer, New York
- Zuur A, Ieno EN, Walker N, Saveliev AA, Smith GM (2009) Mixed effects models and extensions in ecology with R. Springer, New York

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

