36. Reporting statistics

In this book, we have focused on the analysis and visualisation of different kinds of data. This chapter focuses on how to report these analyses in scientific writing, which might include practical write-ups, reports, dissertations, or peer-reviewed manuscripts. What follows are some general guidelines for the biological and environmental sciences. These guidelines are not necessarily hard rules, but following them will improve the clarity with which statistical analyses are communicated.

36.1. Before collecting data

When conducting a study that requires statistical analysis, it is important to plan ahead when setting up experiments or field observations and collecting data. Without proper planning, there is a risk that the collected data will be impossible to analyse or inappropriate for the scientific hypothesis of interest. Finding out an analysis will not work after spending a considerable amount of time on data collection can be frustrating and demoralising, especially when it could have been avoided by more careful planning. Before collecting any data, it is therefore important to plan, or even simulate, what a dataset will look like in a tidy format. This can be done by creating a spreadsheet with appropriate data columns prior to data collection. It is important to think about what kinds of analyses will be performed and how large of a sample size might be needed. It is often a good idea to conduct a power analysis before data collection (Steidl et al., 1997). A power analysis is a formal way of estimating how big of a sample size is necessary to reject the null hypothesis given a particular effect size (Jones et al., 2003). For example, suppose that we want test whether 2 species of fig wasps have different mean ovipositor lengths, as in the example from Chapter 34.3. A power analysis might be used to determine how many fig wasps need to be collected and measured to detect that a given difference in mean ovipositor length (e.g., 1.0 mm) is statistically significant (i.e., P < 0.05). In other words, if the difference in mean ovipositor length between two species is 1.0 mm, then how big of a sample size is needed to reliably reject the null hypothesis (P < 0.05) when running a t-test? Ideally, such a power analysis could be informed by preliminary data. This is not always possible, and an introduction to power analysis is beyond the scope of this book. What is most important is that the statistical analysis is planned as much as possible *before* data are collected. Failing to plan can result in a complete inability to run an effective statistical analysis. This happens when data are not collected in a way that satisfies the assumptions of statistical tests, or if a sample

size is grossly insufficient. Whenever in doubt, it is strongly recommended to consult with a statistician prior to data collection.

36.2. Statistical reporting

Good scientific reporting requires a clear communication of how data were collected and analysed. Guidelines for explaining data analysis can vary across different fields of study. What follows are recommendations for communicating data analyses in the biological sciences, as might be applied to writing reports, dissertations, or scientific manuscripts. Such documents are typically (though not always) divided into abstract, introduction, methods, results, and discussion sections.

36.2.1. Abstract

The abstract is a brief summary of a scientific paper, and should be a self-contained paragraph of somewhere between 150-500 words that provide some relevant background and communicate the most important findings of the paper. A brief explanation of the methodology is important, but there is usually no need to state what statistical tests were performed. It is also not usually strictly necessary to present statistical output in the abstract, although doing so can often be useful (Andrade, 2011).

36.2.2. Introduction

The Introduction of a scientific report should provide enough background for the reader to understand the research question and its importance. Introductions should start by broadly establishing the field of study and its importance, then provide a summary of relevant work that has been previously conducted, and finally explain the knowledge gap that the report will address (Woodford, 1999; Turbek et al., 2016). In other words, the Introduction should ideally finish by explaining what is not yet known, but will be addressed by the scientific report. No statistical methods or analyses need to be presented in the Introduction because the scientific question needs to be introduced before the methods or statistics will make sense (Bouma, 2000). Nevertheless, it is important to clearly state the scientific question so that readers will be able to appreciate why particular statistical techniques are being introduced in the Methods.

36.2.3. Methods

The Methods section should explain how data were collected and analysed, including a justification of research design and analysis choice if appropriate (Bouma, 2000). Before explaining the statistical analyses used, it is important to explain what data were

collected and how. This includes details about biological species, field locations, chemicals and equipment used, experimental procedures, and ethics (of Science Editors, 2006; Woodford, 1999). Units of measurement should also be provided for all variables (Lang and Secic, 1997). The Methods should include the name of any statistical tests used (of Science Editors, 2006), and whether or not data conform to test assumptions (or, e.g., if data need to be transformed to fit the assumptions of statistical tests). The statistical software used to analyse the data, such as Jamovi (The Jamovi Project, 2022) or R (R Core Team, 2022), should also be cited (Lang and Secic, 1997; of Science Editors, 2006). For example:

> "Because stigmatic pollen counts were not normally distributed, they were analysed using non-parametric statistics (Mann-Whitney U test) for pairwise comparisons among four times of collection at the 0.05 level of significance. All data analysis was conducted in SPSS, version 17.0. All means are presented with standard errors ($\pm$ SE)" (Xiong et al., 2013).

> "All statistical analyses were carried out using R 3.1.0 (R Core Team 2012). […] ANOVAs were carried out to test whether the three soil-conditioning treatments affected plant biomass and trait and soil N availability, and whether there was an effect of drought, with a two-way interaction term between soil conditioning and drought. Non-constancy of variances was evaluated using Levene's test in the car package of R (Levene 1960), and normality was ascertained and corrected for where necessary using Box–Cox transformations in the MASS package (Box and Cox 1964)" (Fry et al., 2018).

Often, this statistical analysis is reported in a separate subsection of the methods. It is not necessary to state null hypotheses for statistical tests (Lang and Secic, 1997). Overall, the Methods should explain the design of the study, data collection, and statistical analysis in sufficient detail for readers to be able to evaluate and potentially repeat the study (Woodford, 1999). To avoid length Methods sections, methodological and statistical details can sometimes be placed in supplemental material (i.e., a separate document to accompany the main text).

36.2.4. Results

The Results section should report the results of statistical tests, usually without any scientific interpretation. In other words, the primary focus should be on the patterns observed in the data, not an explanation of their underlying biological or environmental cause. This allows readers to evaluate results objectively, without interpretation from the author.

Results should report relevant statistics and their uncertainty. Confidence intervals should be provided for key summary statistics (Lang and Secic, 1997). Key statistics from statistical tests can be reported in parentheses. For example, we might state, "an

Independent Samples t-test found a significant difference between mean fig wasp ovipositor lengths (t = 2.419, df = 30, P = 0.022)". Relevant statistics for specific tests are explained below in section 36.4. In general, it is good to report a p-value exactly, but if the value is less than 0.001, it is fine to report as $P < 0.001$ (of Science Editors, 2006). Direct measurements should be reported to the level of precision allowed by the measuring device used, and summary statistics such as means and standard deviations should be reported to just 1 significant figure higher than the raw measurements (of Science Editors, 2006). Note that this only applies to *reporting* statistics. For the actual statistical analysis, all digits should be used (Lang and Secic, 1997).

36.2.5. Discussion

The Discussion section should present the main conclusions of the study, followed by any secondary conclusions (Woodford, 1999). It interprets the results and explains their importance in the broader context of the scientific literature (Turbek et al., 2016). The Discussion should also critically evaluate the strengths and weakness of the report and attempt to generalise its conclusions (Turbek et al., 2016). The open-ended nature of the Discussion section often makes it the most difficult section to write for many researchers.

36.3. Figures and tables

Figures and tables should be used whenever possible to efficiently communicate results. It is best to avoid presenting the same information in figures and tables, and there is no reason to duplicate information in the main text of a manuscript if it is presented clearly in a table or figure (Lang and Secic, 1997). All figures and tables should be referenced within the main text. This is usually done parenthetically (see below for examples). Figures and tables should always include an informative caption that is self-contained so that readers can understand the information being presented in the figure or table without referring back to the main text.

36.3.1. Figures

Figures can be used to present data and statistical results visually. Figure axes should always be clearly labelled with the appropriate units of measurement. Captions should be placed below figures. Colour should not be used unless it is necessary. When using colour in figures, it is important to use accessible colour combinations and high contrast (Jambor et al., 2021; Painter et al., 2021). For electronic documents, alt images should be provided whenever possible.

Different kinds of figures are appropriate for different kinds of data (see Chapter 10 and Chapter 29.1). Histograms and box-whisker plots are most effective for presenting the distribution of continuous data (Lang and Secic, 1997; Weissgerber et al., 2015). Barplots and pie charts present categorical data (Lang and Secic, 1997; Weissgerber et al., 2015). Scatterplots are used to compare quantitative variables.

When presenting a figure with multiple panels, it is important to distinguish between panels, usually with letters (e.g., A, B, C). Examples of figures can be found throughout this book.

36.3.2. Tables

Tables can be used in multiple ways within scientific reports, including to present data, provide descriptions of variables, and report statistical output. Tables should include informative column headings, and units of measurement should be reported in headings or in the table caption. Table captions should be placed above tables. Some statistical analyses are usefully reported in the form of a table, such as counts in contingency tables (e.g., for a χ^2 test of association), or the results of an ANOVA or linear regression. Jamovi output tables are typically appropriate for reporting statistical tests, and these can be copied and pasted into documents (The Jamovi Project, 2022).

36.4. Statistical tests

There is no universally accepted way to report the results of statistical tests, but some suggestions follow for reporting the output of t-tests, ANOVAs, Mann-Whitney tests, Wilcoxon signed-rank tests, correlation coefficients, and linear regression.

36.4.1. Reporting t-tests

For t-tests, it is useful to report the test statistic (t-score), degrees of freedom (df), and the p-value. or example, for a test with a df = 14, t-statistic of 9.56, and p-value of 0.02, we might write that "two groups different significantly from each other ($t_{(14)} = 9.56$, $P = 0.02$). Note that the subscript of t ($t_{(14)}$) is used to indicate the 14 degrees of freedom.

36.4.2. Reporting ANOVA

For a one-way ANOVA, it is a good idea to report the F value, the p-value, and the degrees of freedom for the effect and error terms. For example, for an ANOVA with an F-value of 11.37, p-value of 0.0002, and with 2 degrees of freedom for the effect term (i.e., 3 groups) and 25 degrees of freedom for error term, we could write that "plant height

differed significantly as a function of soil type ($F_{(2,25)} = 11.37$, $P < 0.001$)". Note that because the actual p-value ($P = 0.0002$) is so low, we can simply write $P < 0.001$.

A two-way ANOVA reports the same values, but should include all of the main effects and interactions. An example might be, "the main effect of soil type was significant ($F_{(2,24)}$, $P = 0.002$), as was the main effect of plant species ($F_{(3,25)} = 12.70$, $P < 0.001$). The interaction of these two factors was not significant ($F_{6,25} = 1.71$, $P = 0.160$)". Note that this information might also be presented in the form of an ANOVA table, in which case it is not necessary to include all of these statistics in the main text.

36.4.3. Reporting a Main-Whitney U test

When reporting the results of a Mann-Whitney test, the W statistic (sometimes indicated instead with a Z) and the p-value should be reported. For a test between two groups in which W = 1596.5 and the p-value = 0.002, an example would be, "the total length of sparrows differed significantly with the survival status of sparrows ($W = 1596.5$, $P = 0.002$)".

36.4.4. Reporting a Wilcoxon signed-rank test

When reporting the results of a Wilcoxon signed-rank test, it is good to include the W statistic and p-value as with the Mann-Whitney test. An example would be, "photosynthetic rate significantly in plants before and after frost ($W = 14.01$, $P = 0.007$)".

36.4.5. Reporting Chi-square tests

To report results from a χ^2 goodness of fit test, it is best to include the χ^2 test statistic, degrees of freedom, and the p-value. For example, for a χ^2 value of 1.2, degrees of freedom of 2, and p-value of 0.549, we could write, "there is no significant difference between expected and observed counts of chosen dam sizes ($\chi^2_{(2)} = 1.2$, $P = 0.549$)".

Reporting results from a χ^2 test of association can be done similarly. For example, we might observe a χ^2 value of 4.89, 2 degrees of freedom for the first categorical variable and 1 for the second categorical variable, and a p-value of 0.549. We could report this as, "there is no association between dam size and operating system ($\chi^2_{(2,1)} = 4.89$, $P = 0.549$)" (see Chapter 28).

36.4.6. Reporting correlation coefficients

To present the results of a test of the correlation coefficient (either the Pearson or Spearman Rank correlation coefficient), we can report the correlation coefficient and the corresponding p-value. An example of this is, "egg production was positively correlated with body mass ($r = 0.898$, $P = 0.038$)".

36.4.7. Reporting regressions

When reporting the results of a regression, it is important to report the R-squared value, regression coefficients, and p-values of model or regression coefficients. It is also advised to report the equation of the linear regression (Lang and Secic, 1997). There are multiple ways of presenting this information clearly (in some cases, it might make sense to present the equation in a figure rather than the text). One example would be, "sparrow body mass (g) increased significantly with sparrow total length (mm; $R^2 = 0.341$, $b_1 = 0.242$, $P < 0.001$) according to the model, $mass = -13.07 + 0.242(length)$". When reporting multiple regression coefficients, relevant subscripts can be added for each coefficient (e.g., b_1, b_2, or more explicitly b_{mass}, $b_{wing\,length}$). For large multiple regressions with many coefficients, it might make sense to report results in the form of a table.

36.5. Conclusions

This chapter has presented some suggestions for reporting data and statistical output. These suggestions can be applied to scientific reports, dissertations, or scientific manuscripts. The recommendations in this chapter should mostly be treated as guidelines. There is no universally agreed upon way of reporting statistical output, but the guidelines of this chapter should be sufficient unless instructed otherwise. Lang and Secic (1997) provides a more comprehensive guide to statistical reporting.