2

One-Sample Hypotheses

As biologists, we might be interested in making inferences about populations based on samples.[1] We collect data from a subset of individuals and then use statistical methods to draw conclusions about the broader population from which the sample was drawn. One fundamental type of inference we often make is whether the population mean (or median) is equal to a specific, pre-determined value. This is where *one-sample hypothesis tests* come into play. They allow us to address questions such as:

- Does the average height of *Aloe dichotoma* (quiver trees) in a newly protected area in the Richtersveld differ from the known average height of this species across its range?
- Is the median concentration of lead in the soil near a mining operation in Limpopo above a regulatory threshold?
- Has the mean number of African wild dog pups per litter in the Kruger National Park changed after a new management strategy was implemented?

These questions are not just academic. They have real-world implications for conservation efforts, environmental management, and our understanding of ecological and evolutionary processes in unique ecosystems, not only in South Africa as the examples would suggest, but anywhere in the world.

To reiterate, one-sample hypothesis tests are statistical procedures designed to assess whether a population parameter (typically the mean or median) is likely to be different from a hypothesised value. The distinction between mean and median is important. The mean is interesting when we know our sample data are normally distributed (Section 2.1), and if this is the case, it calls for a *Student's* or *Welch's* t-*test*. If, however, our data are non-normal, the median is the measure of central tendency that must be tested with a *Wilcoxon signed-rank test*.

Student's or Welch's t *test*
Wilcoxon signed-rank test

We might use μ_0 to represent the *hypothesised value* of the population mean in a one-sample hypothesis test. As an example, let's say we want to test the hypothesis

1. Inference in statistical hypothesis testing is the process of drawing conclusions about an underlying population from sample data, typically by quantifying how surprising observed outcomes would be under a proposed null hypothesis. It allows us to assess whether any detected differences or relationships are genuinely present or merely due to random variation.

BOX 2.1 NULL AND ALTERNATIVE HYPOTHESES

In the example of the Cape Sugarbirds, we want to assess if the population mean, $\bar{x}$, is the same or different from the hypothesised value, $\mu_0 = 15$. The two types of hypotheses are therefore:

- $H_0 : \bar{x} = \mu_0$, or $H_0 : \bar{x} = 15$
- $H_a : \bar{x} \neq \mu_0$, or $H_a : \bar{x} \neq 15$

This is the same as saying:

- $H_0 : \bar{x} - \mu_0 = 0$
- $H_a : \bar{x} - \mu_0 \neq 0$

Note that in this formulation of the hypotheses, we do not make a distinction between whether we expect $\bar{x}$ to be $<$ or $>$ μ_0, calling for a two-sided one-sample *t*-test.

that the mean beak length of Cape Sugarbirds in a particular fynbos region is 15 mm, we would say that our hypothesised value is $\mu_0 = 15$. The goal is to determine whether the evidence from our sample suggests that the true population mean/median is consistent with the null hypothesis or if it deviates significantly from it. The *null hypothesis* states that there is no difference between the true population mean and the hypothesised value. For instance, if we were interested in testing whether the mean beak length of Cape Sugarbirds is different from 15 mm, we would have a *two-tailed hypothesis*. Alternatively, if we were interested in knowing if the mean beak length is *either* below *or* above 15, we would have a *one-tailed hypothesis*.

One- and two-tailed hypotheses.

It is important to make a distinction at this point. Although we are ultimately interested in the population we are sampling, our calculations (i.e., the statistical test) are done on our *sample* data. For example, we calculate the mean of our sample, which we refer to as $\bar{x}$. We then use $\bar{x}$ to make inferences about the population.

Null and alternative hypotheses.

The process involves formulating a *null hypothesis* (often denoted as H_0). This hypothesis represents the scenario where the population mean/median equals the hypothesised value ("is no different from," i.e., *null*), and an *alternative hypothesis* (often denoted as H_1 or H_a), which represents the scenario where the population mean/median differs from the hypothesised value (either greater than, smaller than, or either greater or smaller than) (Box 2.1).

We then collect data, calculate a *test statistic* (e.g., a *t*-statistic in the case of a one-sample *t*-test, or a *z*-statistic for a one-sample *z*-test – see the examples below), and determine the probability of observing our sample data (or something more extreme) if the null hypothesis were true. This probability is known as the *p*-value. If the *p*-value is very small (typically below a pre-determined significance level, often 0.05), we reject the null hypothesis in favour of the alternative hypothesis, suggesting that the population mean/median is likely different from the hypothesised value.

In this chapter, we will explore three common one-sample hypothesis tests:

1. **One-sample *t*-test** (**?@sec-one-sample-t-test**): Used to compare the mean of

a normally distributed population to a hypothesised value when the population standard deviation is unknown. It's a versatile test applicable in many biological contexts.
2. **One-sample *z*-test** (**?@sec-one-sample-z-test**): Applied when you want to compare the mean of a normally distributed population to a hypothesised value and you know the population standard deviation. While less common in practice, it's conceptually important.
3. **One-sample Wilcoxon Signed-Rank Test** (**?@sec-one-sample-wilcoxon-test**): A non-parametric test used to compare the median of a population to a hypothesised value. It's valuable when dealing with data that don't meet the assumptions of normality.

Throughout this chapter, we'll delve into the assumptions, calculations, and interpretations of these tests. We'll also illustrate their application using real biological examples from South African contexts and demonstrate how to perform these tests in R. My aim is not just to teach you the mechanics of these tests but also to develop a deeper understanding of their underlying principles and their role in making sound, data-driven biological inferences. By the end of this chapter, you will have the knowledge and skills needed to apply these one-sample hypothesis tests in your research and confidently interpret the results, contributing to the conservation and understanding of South Africa's incredible biodiversity.

2.1 NATURE OF THE DATA AND ASSUMPTIONS

One-sample *t*-tests require data of a specific nature. The appropriateness of a one-sample *t*-test hinges on several key assumptions about the data you've collected.

Let's consider what these data requirements mean:

1. **Data Type:** A one-sample *t*-test is intended for *quantitative data* that are measured on a *continuous* or *interval/ratio scale*.
    - *Continuous variables* can take any value within a given range. For instance, the height of *Aloe dichotoma* in our example is a continuous variable, as it can theoretically take on any value within a reasonable range (e.g., 2.5 meters, 3.82 meters, 5.175 meters).
    - *Interval/ratio variables* are a type of continuous data where the differences between values are meaningful, and in the case of ratio data, there is a true zero point. Beak length of Cape Sugarbirds, for instance, is a ratio variable because a beak of 4 mm is twice as long as a beak of 2 mm, and a beak length of 0 mm implies the absence of a beak (a biological impossibility, of course, but it illustrates the concept of a true zero).

Importantly, categorical or ordinal data are not suitable for a one-sample *t*-test. For example, if you were to categorise the soil contamination levels near a mine in Limpopo as "low," "medium," or "high," you would not be able to use a one-sample *t*-test to compare these categories to a specific threshold.

2. **Independence:** The observations in your sample must be *independent* of each other. This means that the value of one observation should not influence or be related to the value of any other observation.

   - In our African wild dog example, the number of pups in one litter should not affect the number of pups in another litter. If, for instance, you were sampling litters from the same pack over multiple years, the data points might not be truly independent because they may share certain characteristics from the same pack or the same matriarch, which will be important to take into account.
   - Similarly, when measuring quiver tree heights, each tree measured should be selected independently. If you were to measure all the trees within a small, isolated clump, they might be genetically related or influenced by the same micro-environmental conditions, potentially violating the independence assumption.
3. **Normality:** A one-sample t-test assumes that the data (or more precisely, the sampling distribution of the sample mean) are approximately *normally distributed*. The normal distribution, also known as the Gaussian distribution, is a bell-shaped curve that is symmetric around the mean.
   - The raw data themselves don't necessarily need to be perfectly normally distributed, especially if the sample size is large. The Central Limit Theorem tells us that the distribution of sample means will tend to be normal, even if the population distribution is not, as long as the sample size is sufficiently large (often considered to be around 30 or more).
   - However, if your sample size is small and the data deviate strongly from normality (e.g., heavily skewed or with many outliers), a one-sample t-test might not be appropriate. In such cases, a non-parametric alternative like the Wilcoxon Signed-Rank test (which we will discuss later) might be more suitable.
4. **Random Sampling:** Ideally, your data should come from a *random sample* of the population of interest. Random sampling helps ensure that your sample is representative of the broader population, allowing you to make valid inferences.
   - For instance, if you're studying lead contamination near a mine, you should randomly select sampling sites within the area of interest, rather than, for example, only sampling sites where it is convenient and accessible to do so.
   - In practice, truly random sampling can be challenging to achieve in biological studies. However, we must strive to apply sampling methods that minimise bias and maximise the representativeness of your sample.

Before conducting a one-sample t-test, you must assess whether your data meet these requirements. Visual inspection of the data (e.g., using histograms or Q-Q plots), along with formal statistical tests for normality (e.g., the Shapiro-Wilk test), can help you determine if a one-sample t-test is appropriate, or if you'd be better served by a Wilcoxon signed-rank test. We will explore these diagnostic techniques in R later in the chapter. You can ensure that your statistical analysis is valid and that your conclusions are robust and meaningful by exploring the nature of your data before you commit to your hypothesis testing.

2.2 R FUNCTION

One- and two-sided Student's and Welch's *t* tests can be applied with the `t.test()` function. The help file for the function, accessed via `?t.test` provides explicit use cases and examples. When you have non-normal data, you can use the `wilcox.test()` for the Wilcoxon signed-rank test. Examples of both function are shown below.

2.3 EXAMPLE: CRAB BODY TEMPERATURES

The data for the first example are drawn from Jerold H. Zar's excellent book, *Biostatistical Analysis* (2010, 5th ed., p. 98). These data consist of measurements of the body temperatures of 25 crabs. Prior to measurement, the crabs were exposed to air at 24.3°C for an unspecified duration. The question posed is whether, on average, the crabs' body temperatures match the air temperature to which they were exposed.

2.3.1 *Do an Exploratory Data Analysis (EDA)*

To explore the data, we will calculate the mean ± SD, create a hisotogram for a visual indication of the data's normality (one of the assumptions for Student's or Welch's *t*-tests), and do a Shapiro-Wilk test as a statistical measure for normality. We will not test for homoscedasticity in this example because we are only doing a one-sample *t*-test.

```
crabs <- c(25.8, 24.6, 26.1, 22.9, 25.1, 27.3, 24.0, 24.5, 23.9,
           26.2, 24.3, 24.6, 23.3, 25.5, 28.1, 24.8, 23.5, 26.3,
           25.4, 25.5, 23.9, 27.0, 24.8, 22.9, 25.4)

summary(crabs)
>    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
>  22.900  24.000  24.800  25.028  25.800  28.100
mean(crabs)
> [1] 25.028
var(crabs)
> [1] 1.800433
sd(crabs)
> [1] 1.341802
```

```
# Make a histogram
ggplot(data = data.frame(crabs), aes(x = crabs)) +
  geom_histogram(binwidth = 1.0, fill = "skyblue",
                 color = "black") +
  labs(title = "Crab Body Temperatures",
       x = "Body Temperature (°C)", y = "Frequency") +
  theme_bw()
```

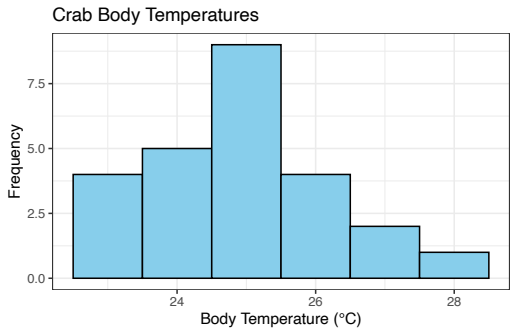

FIGURE 2.1. Histogram of the crab dataset.

```
# Shapiro-Wilk test for normality
# H0: The data are normally distributed
# Ha: The data are not normally distributed
# p-value > 0.05 indicates normality
# p-value < 0.05 indicates non-normality

shapiro.test(crabs)
>
>   Shapiro-Wilk normality test
>
> data:  crabs
> W = 0.97619, p-value = 0.8008
```

Looking at the summary statistics, the mean body temperature of the crabs is 25.1°C, with a standard deviation of 1.3°C. The histogram in Figure 2.1 shows a roughly symmetric distribution, which is a good indication that the data might be normally distributed. The Shapiro-Wilk test for normality yields a p-value of > 0.05, suggesting that the data are consistent with a normal distribution. Given these results, we can proceed with a one-sample t-test to compare the mean body temperature of the crabs to the air temperature of 24.3°C.

2.3.2 *State the Hypotheses*

The null and alternative hypotheses for this one-sample t-test are:

- H_0: The mean body temperature of the crabs is 24.3°C.
- H_a: The mean body temperature of the crabs is not 24.3°C.

This is a two-sided test because we are interested in determining whether the mean body temperature is different from the hypothesised value of 24.3°C, without specifying whether it is greater or less than this value.

2.3.3 *Apply the One-Sample* t*-Test*

```
t.test(crabs, mu = 24.3)
>
>   One Sample t-test
>
> data:  crabs
> t = 2.7128, df = 24, p-value = 0.01215
> alternative hypothesis: true mean is not equal to 24.3
> 95 percent confidence interval:
>  24.47413 25.58187
> sample estimates:
> mean of x
>    25.028
```

The results of the one-sample t-test indicate that the mean body temperature of the crabs is significantly different from the air temperature of 24.3°C ($t(24) = 2.713$, $p < 0.05$). The 95% confidence interval for the mean difference in body temperature is 24.5°C to 25.6°C. This suggests that the crabs' body temperatures, on average, are not equal to the air temperature to which they were exposed.

2.3.4 *Reporting*

In a scientific report or publication, you would present the results of the one-sample t-test in a concise manner. We also need to provide enough information for readers to understand the analysis and its implications, and that already starts with the methods. Here follows an example of how you might report the methods section (a hypothetical example provided), and results of the one-sample t-test for the crab body temperature data.

Methods

The study investigated the body temperatures of 25 *Ovalipes trimaculatus* (three-spot swimming crabs) to determine whether their mean body temperature differed significantly from the ambient air temperature to which they were exposed. The crabs were collected from the intertidal zone within the West Coast National Park, South Africa, during mid-summer (February). Only adult crabs were used in the experiment, with individuals ranging in carapace width from 10 to 14 cm (mean ± SD: 12.6 ± 1.1 cm). Both male and female crabs were included (13 males, 12 females). Crabs were visually inspected prior to the experiment to ensure they were free from injuries, signs of molting, or other physical abnormalities that could influence their physiological responses.

The experiment was conducted in a controlled laboratory setting. Crabs were individually housed in open containers within a climate-controlled chamber set to 24.3°C, which reflected the average air temperature of the intertidal zone during the sampling period. The containers allowed for unrestricted airflow around the crabs. To ensure thermal acclimation, crabs were exposed to the experimental air temperature

for 3 hours prior to body temperature measurements. During this period, no food or water was provided to ensure standardised conditions across all individuals. Body temperature measurements were taken using a calibrated digital thermocouple probe inserted into the ventral body cavity of each crab to a standardised depth of 2 cm. All measurements were conducted by the same researcher to minimise variability in technique.

The mean body temperature of the 25 crabs was calculated, and a one-sample t-test was performed to evaluate whether the mean body temperature significantly differed from the ambient air temperature of 24.3°C. Assumptions of the one-sample t-test were assessed prior to analysis. Normality of the data was checked visually using a histogram and formally tested using the Shapiro-Wilk test. Independence of observations was ensured by housing crabs individually and randomising the order in which body temperatures were measured. All statistical analyses were performed in R (v4.2.1), and significance was assessed at an alpha level of 0.05.

Results

The mean body temperature of the crabs was 25.1°C (SD = 1.3°C) after a 3-hour exposure to air. A Shapiro-Wilk test for normality confirmed that the data were consistent with a normal distribution ($p > 0.05$), which satisfied the normality assumption required for the one-sample t-test. The t-test demonstrated that the mean body temperature of the crabs was significantly different from the air temperature of 24.3°C, $t(24) = 2.713$, $p < 0.05$. The 95% confidence interval for the mean difference in body temperature ranged from 24.5°C to 25.6°C. These results indicate that the crabs' body temperatures, on average, were significantly higher than the ambient air temperature to which they were exposed.

2.4 EXAMPLE: LIMPOPO LEAD CONCENTRATIONS

In this example, we will consider a hypothetical scenario where we are interested in assessing whether the concentration of lead in soil samples collected near a mining operation in Limpopo exceeds a regulatory threshold of 40 mg/kg. The data consist of lead concentrations (in mg/kg) from 100 soil samples collected at random locations near the mine.

```
Pb_conc <- read.csv("data/lead_concentration_limpopo.csv")
summary(Pb_conc)
>  Lead_Concentration_mg_per_kg
>  Min.   :  7.99
>  1st Qu.: 32.83
>  Median : 45.75
>  Mean   : 56.31
>  3rd Qu.: 66.48
>  Max.   :182.84
mean(Pb_conc$Lead_Concentration_mg_per_kg)
> [1] 56.31499
sd(Pb_conc$Lead_Concentration_mg_per_kg)
```

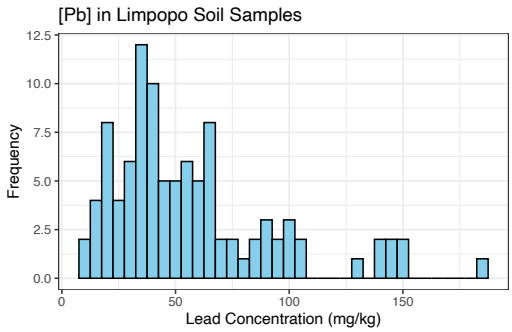

FIGURE 2.2. Histogram of the lead concentration data.

```
> [1] 36.05891
median(Pb_conc$Lead_Concentration_mg_per_kg)
> [1] 45.74969
```

2.4.1 *Do an Exploratory Data Analysis (EDA)*

```
# Make a histogram
ggplot(data = Pb_conc, aes(x = Lead_Concentration_mg_per_kg)) +
  geom_histogram(binwidth = 5, fill = "skyblue",
                 color = "black") +
  labs(title = "[Pb] in Limpopo Soil Samples",
       x = "Lead Concentration (mg/kg)", y = "Frequency") +
  theme_bw()
```

```
# Shapiro-Wilk test for normality
shapiro.test(Pb_conc$Lead_Concentration_mg_per_kg)
>
>   Shapiro-Wilk normality test
>
> data:  Pb_conc$Lead_Concentration_mg_per_kg
> W = 0.8793, p-value = 1.695e-07
```

The summary statistics for the lead concentration data show a mean concentration of 56.3 mg/kg, with a standard deviation of 36.1 mg/kg. The median concentration is 45.8 mg/kg. The histogram in Figure 2.2 suggests that the data are quite left skewed (right-tailed), and the Shapiro-Wilk test for normality yields a p-value of < 0.05, confirming our suspicion that the data are not normally distribution. Given these results, we must proceed with a one-sample Wilcoxon signed-rank test to assess if the median lead concentration in the soil samples is significantly *higher than*

the regulatory threshold of 40 mg/kg.

2.4.2 *State the Hypotheses*

The null and alternative hypotheses for this one-sample, one-sided Wilcoxon signed-rank test are:

- H_0: The median lead concentration in the soil samples is not greater than 40 mg/kg.
- H_a: The median lead concentration in the soil samples is greater than 40 mg/kg.

This is a one-sided test because we are only interested in determining whether the median lead concentration is higher than the regulatory threshold of 40 mg/kg.

2.4.3 *Apply the One-Sample, one-sided Wilcoxon Signed-Rank Test*

```
wilcox.test(Pb_conc$Lead_Concentration_mg_per_kg, mu = 40,
            alternative = "greater")
>
>   Wilcoxon signed rank test with continuity correction
>
> data:  Pb_conc$Lead_Concentration_mg_per_kg
> V = 3542, p-value = 0.000237
> alternative hypothesis: true location is greater than 40
```

We see that the median lead concentration in the soil samples is significantly greater from the regulatory threshold of 40 mg/kg ($V = 3542$, $p < 0.05$). This suggests that the lead concentrations in the soil near the mining operation in Limpopo exceeds the regulatory limit.