Biomedical Teaching Organisation

BIOMEDICAL SCIENCES 2

_____

"Inferring from data"

WORKSHOP HANDBOOK

Name: _____

Key contact:
<mark>Dr Crispin Jordan</mark>

For University of Edinburgh students only:

If you require this document or any of the internal University Of Edinburgh online resources mentioned in this document in an alternative format please contact bms2@ed.ac.uk.

Contents

1. WORKSHOP DESCRIPTION

1.1 Workshop aims:

In this workshop session you will practice key stages the scientific method. You will learn how to formulate a hypothesis, how we might go about collecting data for an experiment, and how to describe and analyze the data.

1.2 Workshop Learning outcomes:

By the end of the workshop, you should be able to
* Understand the difference between a Null and Alternative hypothesis.
* Differentiate between a 'sample' and a 'population'.
* Appreciate the concept of 'statistical significance' but also note controversy that surrounds this concept.
* Explain how statistics uses a 'Null Distribution' to determine whether a result is 'significant'.
* Conduct a Randomization Test to test whether two means differ significantly from one another.

1.3 Workshop structure:

In this workshop, you will address the question,
"Does the average date of observing the first leaf produced ('first leafing') by beech trees differ between Northern vs. Southern populations in the UK?"

In doing so, you will work though many of the steps of the scientific process: you consider the design of an 'experiment', and describe and statistically analyze relevant data.   Hence, this workshop is organized according to **stages of the Scientific Method.**

For each stage of the scientific method, we'll discuss underlying ideas associated with this stage. These fundamental concepts are presented in Boxes, whereas the 'instructions' are not. These concepts are introduced as they arise in your work; so this document is not just a simple recipe for conducting the experiment. **We strongly advise you to read and gain an understanding of the concepts in the boxes, as putting these concepts into practice is the main point of this workshop.**

This document synthesizes much content, and we do not expect you to digest and understand it all fully in this practical. We will re-visit this material in later lectures in this course. Hence, this document partly serves as a text for you to later refer back to when studying the associated subject matter. This partly explains why this document is long.

2. STAGE 1: Formulating a hypothesis

<div style="border: 3px solid magenta; background: #e0e0e0;">

Box 1: Null versus Alternative hypotheses

Before beginning any scientific study, it is crucial to develop a hypothesis, for two reasons. The first is practical: if you do not have a hypothesis in mind, you'll likely not collect data in an effective manner and you will waste your time. The second is more fundamental: all statistical tests are based on testing a hypothesis. So, to meaningfully conduct a statistically test, you need a hypothesis.

Developing a hypothesis is often the most difficult stage of the scientific process. It is easy to develop an unhelpful hypothesis, but studies that truly advance science have great ideas/hypotheses.

A Hypothesis is an explanation for your data; two types of hypothesis exist.

A **Null hypothesis** is one where we propose that whatever factor we wish to study has no effect on our data. For example, we might propose the Null Hypothesis that the day of the week does not affect the probability that a baby is born. In classical statistical approaches we use data to determine whether we're justified in *rejecting* the Null hypothesis. (The meaning of this idea will become clearer later in the workshop.)

An **Alternative hypothesis** is the opposite of a Null hypothesis: it holds that the factor you're studying (Latitude (North vs. South), in our case) does affect your data (date of first leaf observed, in our case). If you conclude that your Null hypothesis is wrong, you accept your Alternate hypothesis. Note that, in practice, scientists usually formulate the Alternative hypothesis in their research (e.g., babies are less likely to be born on weekends than on weekdays: the day of the week affects the probability that a baby is born); however, it is the Null hypothesis that classical statistical approaches test, which is why we mentioned the Null hypothesis first.

</div>

2.1) **Hypothesis.** Following the discussion in Box 1, what would the appropriate Null hypothesis be when comparing the average date at which a beech tree's first leaf is observed between Northern vs. Southern populations in the UK?

Null hypothesis:

Alternative Hypothesis:

3. STAGE 2: Designing an experiment

<div style="border: 3px solid purple;">

Box 2: Who, exactly, do you wish to learn about? Populations versus Samples

Whether or not you require statistics to answer a question depends on what you wish to learn. For example, if you wished to know the difference between the average date at which the first leaf is observed on beech trees between one garden in Scotland and one garden in England, this is a trivial question to answer definitively: you would simply measure the date at which the first leaf is observed for all beech trees in the two gardens (presumably the number of beech trees is small enough to do this easily), calculate the average date for the two gardens, and compare them (ignoring measurement error).

However, in most cases, we are not interested in studying such small groups. Typically, we are interested in the qualities of, say, an entire species (like humans or beech trees). In this case, it is impractical to study all humans / trees; instead, we must study a sub-set of individuals, and we refer to this subset as the '**Sample**'. We then use the sample to make inferences about the larger group we're interested in, which is the '**Population**'. We use statistics to make those inferences about the larger population from the sample.

To illustrate, consider the trivial example above, comparing date of observed first leaf between two gardens. In this case, the individuals in two gardens comprised the entire population. On the other hand, the gardens could be viewed as two samples if we wished to use it to study some larger population (see Box 3 for more thoughts on this).

</div>

In this workshop we will analyze data collected by 'citizen scientists'. Citizen scientists are people who make observations about nature and then share them with a group that organizes the data. For example, citizen scientists might note the day of the year at which they first notice a species of butterfly or bird. With many citizen scientists, the resulting datasets can be massive. This workshop uses data collected by citizen scientists who noted the date at which they noticed the first leaf on beech trees; we refer to this as '**leaf date**' (Julian days). The citizen scientist reported the year, species, and location (e.g., Latitude) of their observations. I obtained the data for this workshop form here: https://datashare.ed.ac.uk/handle/10283/2332. We will use data from the most recent year available (2014) and for beech trees (one of my favourite species). The original dataset was very large (1000's of observations). To create the smaller dataset for our analyses, I determined the average latitude of the observations in the dataset and then labeled all observations below this latitude, South, and all observations greater than average, North. I then used R to randomly selected 100 observations from the North and 100 from the South. Hence, our dataset included 200 observations of the (Julian) date of observing the first leaf (**leaf date**) on a Beech tree (see column, 'day.of.obs')

3.1) **Sample.** The trees in our dataset serve as the sample to test our hypothesis. What population do you wish to make inferences about?

Answer:

4. STAGE 3: 'Collect' data for your experiment

4.1) **Open RStudio and create an RStudio Project.**
   a) Open RStudio on your computer (e.g., to do so on a PC, type 'RStudio' in the search panel; when RStudio appears in a list, click upon the icon).
   b) Use the techniques you learned in the first 'R-Session' to <u>create a project</u> for your work today; provide it with a sensible name (e.g., PracticalRandomization).
   c) Again, using the techniques you learned in the first R-Session, <u>save your script</u> with a name of your choice.
   d) Drag the file, beechV2.csv into the RStudio project folder you just created.

Within RStudio, you should now see the dataset's filename (beechV2.csv) listed in the 'Files' panel.

4.2) **Import the dataset:** We will now **import the dataset into RStudio**:

   a) Begin by typing this generic read.table() command into your script file:

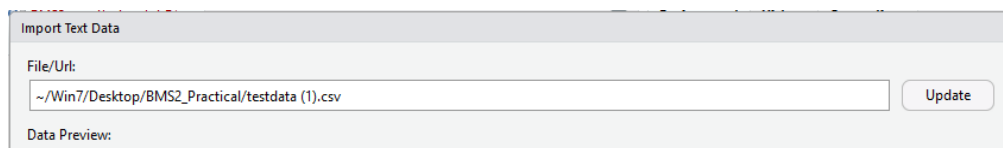   data <- read.table("beechV2.csv ", header=TRUE, sep=',')

   <u>How does this command work</u>? The function read.table() allows RStudio to import data that are stored in a table-like format. In the generic form, above, we provide the function with two pieces of information:
   • The option 'header=TRUE'; this tells RStudio that the first line of the dataset holds the names of the columns.
   • The option, sep=',' tells RStudio that, within rows, the file has a comma between datapoints. (This tells RStudio how to read the file.)

   NOTE also that we chose to save the data in an object called, 'data' [but this object name is arbitrary].

   b) Now, **submit the command** (highlight the code and press 'run').

   c) Look at the data you imported, by typing (and submitting) the object name of the dataframe (e.g., 'data', in the example, above).

   d) If instructions in (a-c) failed to import the data, try the following:
      i) Left-click on the name of your datafile (beechV2.csv)in the Files panel; you should see the option "Import Dataset". Select, "Import Dataset".
      ii) A new window will appear. At the top, you will see the '<u>path</u>' to the dataset and the file's name [in our example here, we've called the

dataset testdata(1).csv]. **Copy this path and file name (i.e., Control-c)**.

iii) Paste (i.e., Control-v) this path and file name inside the inverted commas of the generic command you typed in your script file. For this example, our code looked like this:

```
data <- read.table("~/Win7/Desktop/BMS2_Practical/testdata(1).csv",header=TRUE,sep=',')
```

The code, ~/Win7/Desktop/BMS2_Practical/, is called the '**path**' to your dataset; it tells RStudio where to look for your file (in Win7, on the Desktop, in an RStudio Project folder called, in this case BMS2_Practical). **Note** that that the name of the data file in the example above, testdata(1).csv, differs from the name of our file, beechV2.csv; i.e., to import our dataset, please ensure you use the file name, beechV2.csv.

Submit the read.table() command and look at the imported data (i.e., repeat steps b & c, above, if they previously failed to display the data)

NOTE: You may also import your data by using the Import Dataset option we'd used to identify the file's path. However, this approach can be a bit messier, so it is not explained here.

4.3) **Describe the data**. Because the data will not be sorted with respect to North/South, we will organize the data before we describe it. We will use the function, **order()**, to sort our data by North/South, and we'll place the sorted data into a new dataframe; North/South data are in the column, 'North.South.UK'. For example, if you called your unsorted dataframe, 'data', we can create a new, sorted dataframe, 'data.ordered', with the following command:

data.ordered <- data[order(data$North.South.UK),]

How does this command work? Note that the order() function is placed where the *X* is, here: data[*X*,]. Recall that the square brackets of a dataframe allow us to specify rows and columns within it; the fact that the order() function lies to the left of the comma in data[,] means that we're ordering (i.e., sorting) the **rows** of your dataframe. Within order(), we specified 'data$ North.South.UK': this indicates that we wish to **order the data by information in the column named, North.South.UK**. (Note it is also possible to sort by more than one column; see R's Help for information.) Finally, we created a new object, data.ordered, to hold the newly sorted dataframe.

a) **Find the mean leaf date (day.of.obs) in North and South (UK) beech trees**. Scroll through the sorted dataframe to find which rows contain data from North versus South. (Note that the 'row numbers' on the left will not be

in order.) Now, calculate the mean leaf date (day.of.obs) for beech trees in the North. You can do this by using the function mean() for the appropriate rows and column of your sorted dataset. For example, if day.of.obs for North data were in the **fifth column** on **rows 1 to 4**, you could use the following to determine the mean of those 4 values:

mean(data.ordered[**1:4,5**)]

[NOTE; the above command is an *example*; what command is appropriate for your data?]

Calculated mean leaf date for North UK: _____

Now, use the same approach to calculate the mean date of first leaf for beech trees in South UK.

Calculated mean leaf date for South UK: _____

And, to be complete, calculate the date of first leaf for Northern and Southern populations, together:

Overall mean leaf date (day.of.obs): _____

b) Calculate the difference between the mean North and South leaf dates: *mean North leaf date – mean South leaf date*. You will use this value in a statistical test, below.

Calculated mean North leaf date – mean South leaf date: _____

c) Now, plot two histograms, one of day.of.obs for North trees and and one for South trees, using the techniques you learned in the self-study Session 4.

a. Based on these histograms, do you think that mean leaf date differs between North vs. South??

b. Do the means that you calculated for each region (North / South) seem sensible given your plots?

c. What is the shape of each distribution? For example, are they symmetrical? Describe them in words and sketch them here:

d. Are there any values of 'day.of.obs' that seem suspicious? For example, a day.of.obs that is negative, or greater than 365? You should always check whether unusual data points are real, or are the result of typos when entering the data. Describe suspicious data, if any.

5. STAGE 4: Analyzing your data:

Now that we have calculated the means for each region and visualized the data, we wish to test whether it is likely that average leaf date differs between North and South for our defined **population**. To do this, we need some more concepts.

Box 3: How can we use our sample to test whether mean leaf date of North and South differs in our population (e.g., the UK)? Randomization test

We are interested in whether there is a systematic difference in leaf date between beech trees in the north vs. south UK. However, it is possible for our samples to suggest that the two regions differ in leaf date, just due to random chance. Here's a simple example.

Consider a ***population*** with the following leaf dates of North and South regions:

| North leaf date: | 100 | 92 | 96 | 124 | 107 | 128 | 110 | 113 | 120 | 108 |
| South leaf date: | 109 | 110 | 109 | 128 | 98 | 115 | 112 | 113 | 103 | 101 |

Note that, on average, there is no difference between the North and South leaf dates in this imaginary population (mean = 109.8). However, if you randomly sampled a few (say, 3) of each North and South beech trees from this population, would the North and South averages be very similar? (*Try it!*) If the average leaf dates are basically equal in our pretend **population**, why would the averages of your **samples** not be almost identical, as well?

Answer:

So here's the important question: How can we use our samples to determine whether it is likely that North and South leaf dates are actually different in the larger population (i.e., that (mean North leaf date – mean South leaf date) ≠ 0 in the **population**)? Or, another way of asking this is to say, "Do North and South beech trees actually come from different "populations" that have different mean leaf dates?"

We can address this question by recognizing that, if there really is no difference between North and South beech trees in the population, **then adding the label of 'North' or 'South' to a datapoint is meaningless** (because there is no difference, on average, between North and South beech trees). This means that we can do the following to test whether mean leaf date of North trees differs from that of South trees:

1. Randomize (i.e., randomly shuffle) our measurements with respect to whether they belong to a North versus South;

2. Calculate the difference between North and South average leaf dates for the randomized data;

3. Compare the difference between average North and South leaf date (i.e., mean North leaf date – mean South leaf date) from the randomized data to that from the original (non-randomized) data (i.e., the mean North leaf date– mean South leaf date of the non-randomized data): if they are similar, then we might conclude that North and South average leaf dates do not differ in the true population. On the other hand, if they are quite different, then we might conclude that leaf dates likely differ between the regions in the true population;

Before continuing, absorb what was just outlined in 1-3 of this Box. Ask yourself, "How can we use this to determine whether it is likely that North and South leaf dates are actually different in the larger population?" Keep in mind (hint!) that we have computers at our disposal, so we could perform steps 1-3 many times. (If we did do this, how could we interpret the results?) Think about this on your own for a minute or two, and then. continue reading (cont'd next page)!

Continue when you have finished thinking about 1-3 in this Box...

4. So, how do we determine whether the randomized and non-randomized differences are 'quite different'? Here's how: we do steps 1 and 2 many times to create a **distribution of differences** in mean leaf date from randomized data. This distribution tells us what to expect as a 'typical' difference between North and South when there actually is no difference between North and South mean leaf dates in the true population. We can then compare our result from non-randomized data to this distribution to see how typical (or atypical) our difference is.

We will now use this approach with our data. This is called a 'randomization test'.

We will now follow the arguments in Box 3 to implement a 'randomization test' on our data. We can divvy up the work of randomizing among all the students in the practical, and we'll then work with the results of our shared effort. (Alternatively, we can work with results I prepared for you (diff.csv) following the exact instructions, below.)

5.1 **Randomize the data with respect to region**. We can do this with the following command, placing the randomized data in a new column;

data.ordered$rand1 <- sample(data.ordered[,"day.of.obs"], nrow(data. ordered), replace = FALSE)

How does this command work?
a) We're using the function, **sample()**, which randomly samples data from a column and places it in a new object. The first specification that we made within the sample() function indicates which data will be sampled from; we told R to sample from data.ordered[,"day.of.obs "]: the sorted dataframe, using the column named day.of.obs (note that day.of.obs is on the right of the comma within the [ ]'s). We could have instead used the column number associated with the column, "day.of.obs" (e.g., data.ordered[,5], if "day.of.obs" is in the fifth column), but it is wiser to use the column name to avoid accidentally using the wrong column number.

b) The second entry within sample() tells the function how many numbers to draw. We want to draw as many numbers as there are rows in our dataframe (i.e., we want to shuffle all of the data); we coded this explicitly by using the function **nrow()**, which counts and indicates the number of rows in a dataframe.

c) The third command within sample(), replace = FALSE, tells sample() that, once a number has been randomly selected, it will not be available to select again. For example, if data.ordered[,"day.of.obs "] had 200 numbers to randomly choose from at the start, there would only be 199 numbers to choose from after the first number was randomly selected. After the next

number was selected, only 198 would remain, and so on. This method of random sampling is called "**sampling without replacement**". In contrast, if we had said "replace = TRUE", sample() would pick from a pool of all 200 samples every time.

d) The output of sample() is a vector of randomly selected numbers, which we have placed into a new column of data.ordered, called rand1 (short for "random 1"; but you can call your column something else if you wish). Overall, these commands have the effect of randomly shuffling the numbers in data.ordered[,"day.of.obs "], into a new column of data.ordered, called rand1.

When you perform this command, be certain to use the *ordered* dataframe.

5.2 **Calculate the mean leaf date of North beech trees for this randomized data**. Use the same approach as used above, where you first check which rows contain data from North trees. For example, if North data lie in rows 1 to 17, then the randomized data for North lie here:

data.ordered[1:17,"rand1"]

Again, we've named the column we wish to analyze, as this is the best way to ensure that we use the correct column.

Hence, we can find the mean of these data with:

mean(data.ordered[1:17,"rand1"])

5.3 **Now calculate the mean leaf date of the randomized data for South trees**.

5.4 **Calculate the difference between these means for randomized data, by subtracting the mean for one region from the other** (e.g., Mean North randomized data - Mean South randomized data).

We don't mind which mean is subtracted from which, because we did not hypothesize whether average North leaf dates are *larger* (or smaller) than Soth; we only asked whether they were different from each other (in either direction). (See Box 6 for more on this topic.)   As a result, we can use the absolute value of the difference between means for our analyses because the absolute value only reflects the magnitude of the difference.

5.4 **Record this absolute value of the difference between the mean North and South leaf date for randomized data**. In other words, if the difference you calculated, directly above, was -5.4, then just record 5.4 (i.e., drop the minus sign). If the difference was 3.6, then record 3.6 (the difference was already positive).

Absolute value of the difference between the <u>first </u>pair of means:

_____

5.5 **Repeat steps 1-5 four more times.**

Absolute value of the difference between the <u>second</u> pair of means:

_____

Absolute value of the difference between the <u>third </u>pair of means:

_____

Absolute value of the difference between the <u>fourth </u>pair of means:

_____

Absolute value of the difference between the <u>fifth</u> pair of means:

_____

5.6 You should now have 5 differences between the North and South leaf dates for randomized data. When running this workshop 'in person', each student in the class would have calculated 'differences' exactly like you did and we would pool everyone's data together. When doing this workshop online you can find results from 200 randomization in the file, diff.csv.

5.7 **Import these data** in the file diff.csv as you did with the original dataset. This file contains 200 differences between mean North and South leaf date, based on randomized data.

5.8 **Plot the data in this file using a histogram**, and then **draw it here:**

Box 4: What does this histogram represent?

Think about what this histogram represents. Recall how we made this histogram: we made it by randomly assigning leaf dates to North and South beech trees that were drawn from our original data. Then, because we're interested in whether their means differ, we calculated the difference between the North and South means for these randomized data. As a result, this plot shows **the distribution of differences between mean North and South leaf dates that can arise due to chance, when sampling from a population with no difference between North and South beech tress, on average**. (We can say that there is no difference with certainty because we assigned the measurements randomly with respect to region (''North.South.UK)). This is what is referred to as a **Null Distribution**: more formally, **the Null distribution is the probability distribution of a test statistic when the Null hypothesis is true**; in our case, our test statistic is the difference between the mean leaf date of North and South beech trees (*think: what was our Null hypothesis, and how does it relate the Null distribution?*).

How does the Null distribution help us determine whether average North and South leaf dates likely differ in the **population**? We can compare our measure of this difference from our sample against the null distribution, and ask where our observed difference lies within this distribution. Recall that the Null distribution tells us what kind of difference we might expect to find between mean North and South leaf dates just due to chance, when the Null hypothesis is true (i.e., North and South mean leaf dates do *not* differ in the population). Because we used the absolute value of the random differences, this Null distribution should look, roughly, like a tear-drop, with the highest region on the left, and a long tail (where the height of the distribution shrinks) towards the right. The highest region corresponds to North-South differences (on the x-axis) that occur most commonly due to chance, whereas the lower tail indicates North-South differences that occur more rarely by chance. Hence, if you find the location of your observed difference between North and South average leaf dates on the x-axis, and find that it lies in the tall region of the distribution, corresponding to the values that are most frequently observed, then this suggests that our observed difference is typical of results that could arise by chance when sampling from a population with no average difference between North and South leaf dates (i.e., many North-South differences of this size occur by chance). On the other hand, if you find that our observed difference occurs on the x-axis in the shorter part of the distribution, corresponding to values that are rarely observed by chance (i.e., in distribution's tail), then it is less likely that our observed difference between North and South arose by chance. In other words, this latter case would suggest the difference that we observed between North and South only occurs rarely by chance when the Null hypothesis is true.

5.9 What are the most common values in this histogram?

5.10   Compare your original, observed difference between North and South to the null distribution visually. In other words, find where our observed difference lies in the Null distribution

Does the observed difference lie in one of the ends, or more towards the middle of the Null distribution?

Do you think this means that our observed difference is 'typical' of random differences, or not?

5.11   We'll now compare our observed difference to the Null Distribution in a more precise manner. To do this, we will sort all the values of differences, and we can then read off, where our value lies in relation to these. We can then determine the proportion of values that is above our observed test statistic (observed difference North and South).

Sort the distribution of differences between North and South means based on randomized data, and store the sorted data into a new object. For example, **if** you called the dataframe with these differences, 'diff', and **if** the column that holds the data is called 'diff.abs', you can sort the data like this:

sorted.diff <- sort(diff[,"diff.abs"])

As you might guess, the function, sort(), sorts the data vector in ascending order (this is the default option). *Note, that we used order() for ordering a dataframe, and sort() for ordering a single column in a dataframe (i.e. a vector).*

Now, look at the sorted data. For example, if the sorted data are in an object called, sorted.diff, simply type

sorted.diff

Where in this distribution does your observed difference between North and South leaf dates means lie? Specifically, how may data points lie between the position of your observed difference and the end of the distribution with the highest numbers (i.e., the right-hand end)? Count this number of data points. For example, if your observed difference was 5.2 and the right end of the distribution from the randomizations looked like this:

„„   0.8   1.2   1.6   2.5   3.1   4.2   5.5   5.6

...then two data points would lie between your observed difference and the end of the distribution.

Record the number of data points that lie between the observed difference and the end of the Null distribution: _____

Now, determine how many data points are in the dataset of differences between randomized North and South leaf date means. Recall that the function length() determines the number of observations in an vector. (If you are using the provided file, diff.csv, this file has 200 observations.)

Record the number of data points in the dataset of differences:

_____

Now, we can ask, "what fraction of the Null Distribution is more extreme than my observed value?". For example, if, as illustrated above, we had 2 observations that lay between our observed difference and the end of the distribution, and if there were 225 data points in the Null Distribution, then the proportion of the Null distribution that lies beyond our observed difference equals $2/225 = 0.00889$. This means that only about 0.9% of the Null distribution lies beyond our observed value.

Calculate the proportion of the Null distribution that lies beyond your observed difference. If your observed difference between North and South lies beyond all values in the Null Distribution (i.e., if none of the values in the Null Distribution are greater than your observed difference), then see Box 5.

Record this proportion: _____

Box 5: P-value

The proportion of the Null distribution that lies beyond your observed difference gives the probability of observing the data (or data with an even more extreme difference) due to random chance when the Null hypothesis is true. This is called a **p-value**. How does the p-value help us?

Interpreting p-values

For many decades, standard practice has been to determine whether **less than 5%** of the Null distribution is more extreme than an observed test statistic (i.e., $p < 0.05$). When $p < 0.05$, convention has been to conclude that the observed difference was unlikely to have arisen by chance if the Null hypothesis was true; this situation has been termed, 'statistical significance'. In this case, we would reject the Null hypothesis and we accept the Alternate hypothesis. *(Think: What were our Null and Alternative hypotheses?)* Alternatively, if more than 5% of the Null Distribution is more extreme than our observed result (i.e., $p > 0.05$), then we would fail to reject the Null hypothesis because we lack sufficient evidence to reject it ('statistical non-

significance'). (Note that we do not accept the Null hypothesis – we only fail to reject it.)

Hence, if we followed this tradition to interpret the results in our workshop, we would compare the observed difference between North and South to the Null distribution and can ask whether less than 5% of the Null distribution lies beyond our observed difference in the right-hand tail (i.e., $p < 0.05$).

This tradition is now poised to change. As of 2019, the American Statistical Association advises to abandon this concept of 'statistical significance'. Instead, we should interpret p-values along a 'sliding scale', where small p-values constitute stronger evidence to reject a Null hypothesis and larger p-values provide weaker evidence to reject a Null hypothesis. So, how big or small should a p-value be to provide strong vs. weak evidence to reject a Null hypothesis?

Benjamin et al. (2018; Nature Human Behaviour) provide some insight to interpret p-values. They suggest that p-values around 0.005 (not 0.05) and smaller constitute 'substantial' to 'strong' evidence to reject a Null hypothesis. They also argue that p-values near 0.05 provide relatively weak evidence to reject a Null hypothesis. Given these insights, here as some suggestions:

- p-values around 0.005 provide 'substantial' to 'strong' evidence to reject the Null hypothesis; it follows that $p < 0.005$ provides even stronger evidence to reject the Null hypothesis.

- p-values between 0.005 and 0.05 (approximately) provide 'suggestive' or 'moderate' evidence to reject the Null hypothesis.

- To emphasize this suggestion, immediately above, we can say that p-values around 0.05 provide 'moderate' or 'suggestive' evidence to reject the Null hypothesis.

- p-values notably larger than 0.05 provide weak evidence to reject the Null hypothesis.

To summarize, the proportion of the Null distribution that lies beyond our observed test statistic is called the p-value. In other words, the p-value represents the probability (given that all assumptions of a test are met) of observing the data or getting a more extreme difference than the one we observed when the Null hypothesis is true. Tradition has said that, if this probability is smaller than a specified threshold (usually, 0.05), then we would say that it is unlikely that our observed difference arose by chance, and we reject the Null Hypothesis and accept the Alternate Hypothesis. Statisticians and researchers, including biologists, now suggest that we abandon this tradition and interpret p-values along a sliding scale. Why do statisticians and researchers suggest we abandon the concept of statistical significance? The reasons lie beyond this workshop. But we provide resources, below, for you to learn more if you are interested.

Some technical points on p-values

If an observed test statistic lies beyond **all values** in a Null distribution created by randomization then we might, initially, conclude that the p-value equals zero. However, it is impossible for a p-value to equal zero: i.e., there is *always* some (non-zero) possibility that our test statistic (or a more extreme one) could arise by random chance. In this case how do we calculate a p-value? The answer is that we cannot calculate the p-value (exactly). Instead, we can determine the **range of possible p-values consistent with our results**. For example, consider the file diff.csv: this file contains 200 differences between in mean (randomized) leaf date for North vs. South beech trees. If our observed difference between mean leaf date of North vs. South is greater than all 200 values in diff.csv, then we know our p-value would be **smaller than 1/200**; (1/200 = 0.005) i.e., we know p < 0.005. We would report this inequality in our results.

In today's case we only looked at the proportion of the Null distribution that lay to the right of our observed North-South difference (i.e., we did not consider the proportion to the left). Why did we do this? The answer has to do with the fact that we used the absolute value of differences: i.e., we only allowed differences between North and South to be positive. Therefore, if we wish to determine the proportion of the Null distribution that lies beyond our test statistic, we focus attention to the right of our test statistic. When we do so, we're formally testing whether our observed difference is greater than 95% of the differences that arose by chance when the Null hypothesis is true.

NOTE: Depending on the statistical test that you use in the future, the p-value that you use will derive from either one end or from both ends of the Null distribution that your test uses. In other words, not all statistical tests use p-values from only one end of the Null distribution, as we used today.

More on 'abandoning' statistical significance:
Amrhein et al (2019) 'Scientists rise up against statistical significance': Nature 567: 305-307
Benjamin et al (2018) 'Redefine statistical significance'; Nature Human Behaviour 2: 6-10
Wasserstein et al (2019) 'Moving to a world beyond "p < 0.05"'; The American Statistician 73: 1-19

Some videos:
Crispin Jordan: 'Statistical Significance vs. Effect size' :
https://media.ed.ac.uk/media/Statistical%20Significance%20vs.%20Effect%20Size/1_lpuzgxmi
Crispin Jordan: 'Statistical vs. Biological significance':
https://media.ed.ac.uk/media/Statistical%20vs.%20Biological%20significance/1_0ongbnwd

6. STAGE 6: Interpreting your results:

What do you conclude about your data? Is there evidence to suggest that North and South average leaf dates differ for beech in the UK?

Answer:

NOTE: Even if we conclude that North and South mean leaf dates differ, on average, there is always a chance that our conclusion is wrong. In particular, we must keep in mind that there is always a chance for our results to arise by chance, even if the chance is small. This is not bad – this is simply a part of the scientific process. To be more certain in our conclusions, we must repeat the experiment, or find other supporting evidence in other studies. If this supporting evidence arises, then we gain confidence that we have the correct answer. We must be cautious when interpreting the results from a single study.

7. SUMMARY OF COMMANDS

Function	What it does	Context discussed	Page
ls()	Lists the objects in your workspace	Managing objects	3
rm()	Removes an object from a specified environment	Managing objects	3
help()	Provides explanation of a function named within the ()'s	Getting Help	4
help.search()	Allows search for help given a keyword	Getting Help	4
q()	Quit R	Turning R off	5
c()	Combine values into a vector or list	Creating a vector	7
mode()	Reveals the type of data in an object	Data types	9
as.character()	Converts data to character type	Data types	9
length()	Determines the number of entries in an object	Calculations with vectors	10
max()	Determines the maximum value among data specified	Calculations with vectors	10
min()	Determines the minimum value among data specified	Calculations with vectors	10
sum()	Provides sum of specified data	Calculations with vectors	10
mean()	Determines the mean of specified data	Calculations with vectors	11
read.table()	Reads a table to be imported to R	Getting data into R	16
hist()	Creates a histogram	Plotting data	24
plot()	Creates a scatterplot	Plotting data	26
which()	Command that allows user to choose data "which" meet specified conditions	Choosing subset of data	26
order()	Sorts a dataframe by values in a column	Computer Practical	7
sample()	Draws random samples from a list of specified values	Computer Practical	12
nrow()	Returns the number of rows	Computer Practical	12
sort()	Sorts a vector	Computer Practical	16