This workshop introduces main concepts of hypothesis testing. We will return to these concepts in later lectures.

The "handout" provides written instructions and questions to answer. We will work through this handout together via a series of videos.

Does average 'Leaf date' differ between Northern and Southern populations of Beech in the UK?

Differentiate between a Null and Alternative hypothesis.

Differentiate between a 'sample' and a 'population'.

Consider how to sample appropriate data.

Appreciate the concept of 'statistical significance' and its pitfalls.

Explain how statistics uses a 'Null Distribution'.

Conduct a Randomization Test to evaluate evidence whether two means differ.

'Leaf date': Date in Julian days that first leaf displayed

Data by Citizen Scientists: https://datashare.ed.ac.uk/handle/10283/2332

1) Make an observation (Beech trees open leaves in North and South of UK)
2) Formulate a hypothesis
3) Design experiment (discuss)
4) Conduct the experiment (obtain data)
5) Analyze the data
6) Formulate conclusion

(Synthesize results with other studies, and determine next step)

2) Formulate a hypothesis

Form a Hypothesis

Often most difficult step in science

Essential to design an experiment properly

You must know what you are testing

Null hypothesis:

Posits that the factor we test does NOT affect our data
We test the null hypothesis: sufficient evidence to reject it?

Alternative hypothesis:

Opposite of the Null hypothesis (the factor affects our data)
We normally think in terms of an Alterative hypothesis

Null hypothesis: ????

Alternative hypothesis: ????

3 & 4) Design and conduct experiment

Focus:

Populations vs. Sample

Does average ‘Leaf date’ differ between Northern and Southern populations of Beech in the UK?

How do we answer this question if we’re only interested in knowing about trees in one garden in the North and one garden in the South?

Does average 'Leaf date' differ between Northern and Southern populations of Beech in the UK?

How do we answer this question if we're only interested in knowing about trees in one garden in the North and one garden in the South?

How do we answer this question if we wish to know about…

…All beech trees in UK?

…All beech trees in the world??

Population	Sample

Your class?

What we wish to know about | What we are able to study

Population → Sample: To make study feasible

Population ← Sample: How do we do this? …this is the point of the practical, and statistics

Our sample comes from throughout UK.

What population do you wish to know about?

3) Collect data

(and "design experiment")

If we wished to compare Leaf date between North and South UK beech populations, how should we collect these data (if we did not have them already)?

Describe the data

Read and attempt section 4.3 on your own

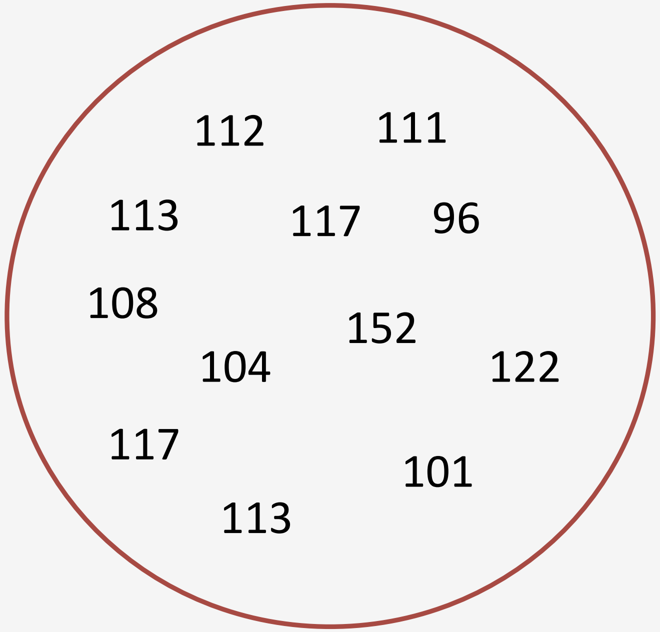
5) Analyze the data

How can we use our sample to evaluate evidence that mean North and South Leaf dates differ?

Population ← Sample

If mean North Leaf date ≠ mean South Leaf date, then we can think of North and South trees as coming from different 'populations'

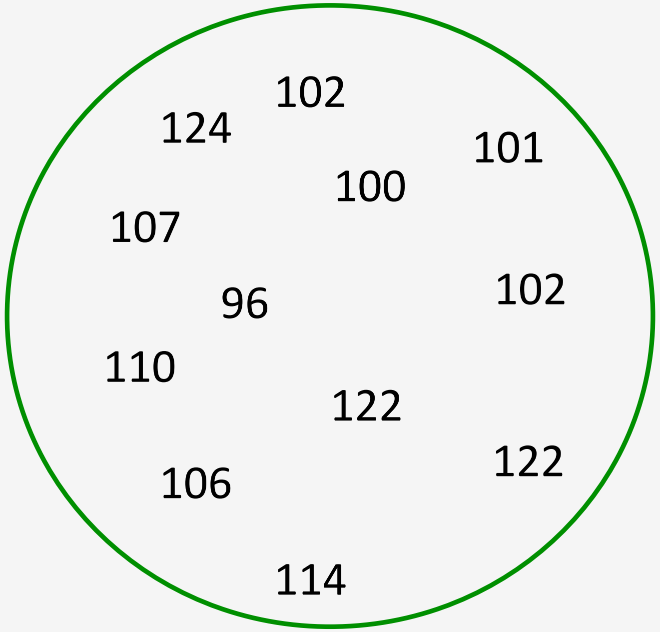
North (1000's)…only 12 shown



Pop. Mean = 114 days

Sampled North = 152, 113, 108, 122
Sample mean = 124

South (1000's)… only 12 shown



Pop. Mean = 109 days

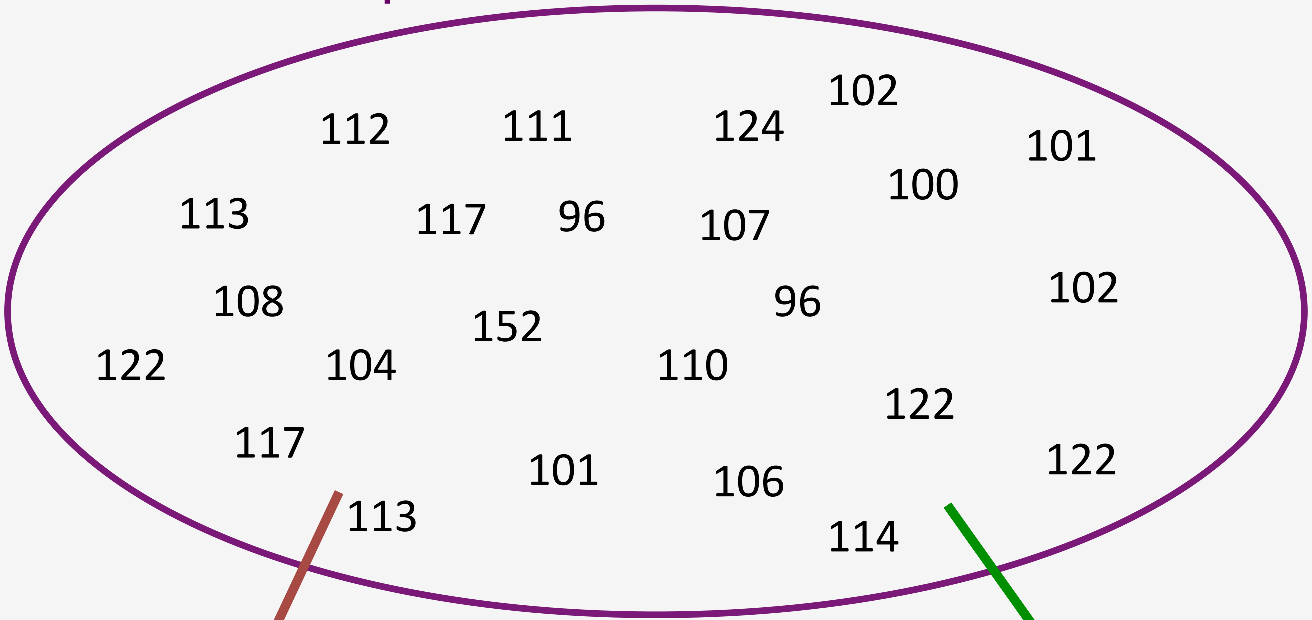
Sampled South = 110, 101, 122, 114
Sample mean = 112

Note: due to random chance, the sample means do not equal the population means

Population ← Sample

But, by random chance, we could likewise obtain what look like different mean Leaf dates for North vs. South when their means are actually the same (i.e., they come from 1 population)

Population of Leaf dates



Pop. Mean = 111 days

By random chance, we select North, whose average Leaf date is greater than that of South:

Sampled North = 152, 113, 108, 122
Sample mean = 124

Sampled South = 110, 101, 122, 114
Sample mean = 112

How can we use our samples to determine whether average North and South Leaf dates likely differ (come from 2 populations of Leaf date) or not (come from 1 population)?

Read, discuss and answer questions in Box 3

Is our observed difference likely to arise by chance if mean Leaf date does not differ for North and South?

i.e., we imagine that North and South are 1 population

Sampled North = 152, 113, 108, 122
Sample mean = 124

Sampled South = 110, 101, 122, 114
Sample mean = 112

Observed difference between sample means: 12

Random Differences:

North = 122 110 108 122 South = 152 113 114 101 Difference = 4.5

North= 110 152 122 108 South = 113 101 122 114 Difference = 10.5

North= 122 110 122 152 South = 114 113 108 101 Difference = 17.5

North = 108 113 152 122 South = 122 110 101 114 Difference = 12

Etc…

How does the observed difference compare to the random differences?

Do 5 times: Randomize your data, and calculate the difference between the (random) North and South Leaf dates

If we had 40 students we would have 40 * 5 = 200 differences between North and South, based on randomized data.

Import the collective random data

Make a histogram of the random data

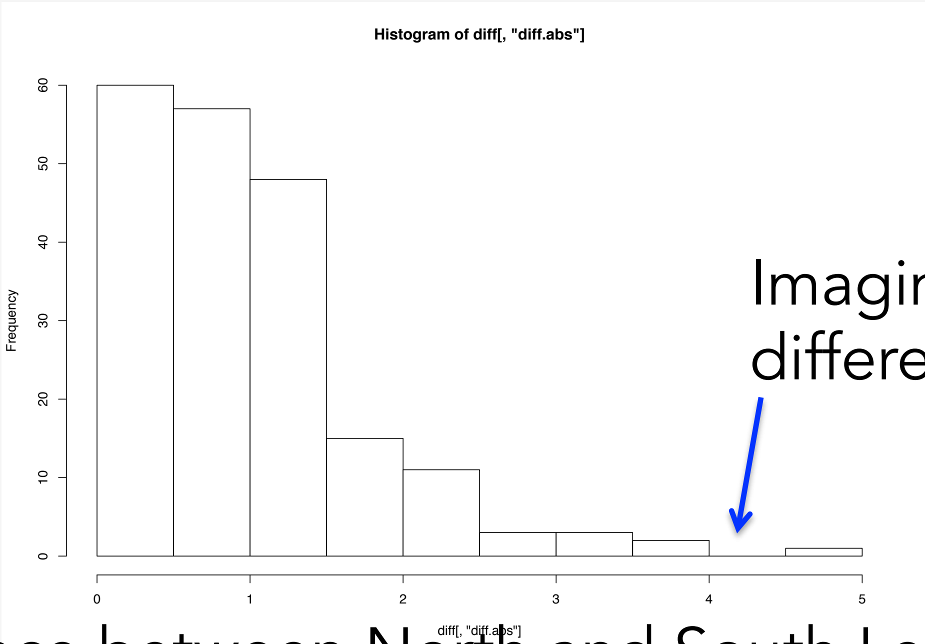
What does this histogram show?

What does this histogram represent?

It is known as a “**Null Distribution**”

Q: why would it be called a Null distribution?

We have 200 observations



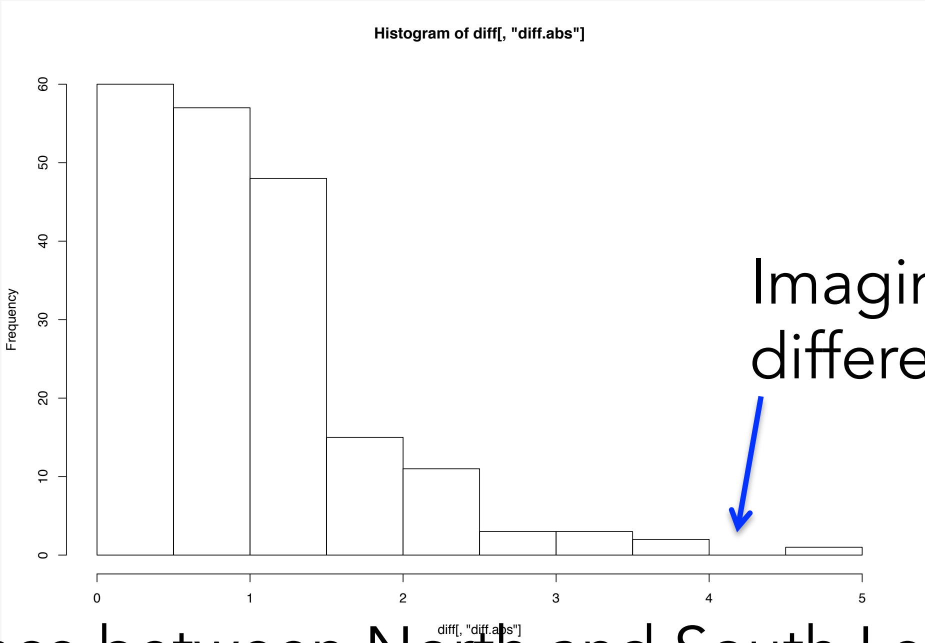
Imagine our observed difference was here

Difference between North and South Leaf dates

Small *Big*

Would an outcome like this suggest that our observed difference was likely to arise by chance?

We have 200 observations



Imagine our observed difference was here

Difference between North and South Leaf dates

Small *Big*

No: this outcome suggests that, if North and South have equal average Leaf dates, we're unlikely to get our observed outcome by chance

Suggests a better explanation: North and South average Leaf dates are not equal; ***we consider rejecting the Null hypothesis***

Describe the Null Distribution

1) Answer Q's 5.9 and 5.10 after Box 4

How can we decide, more objectively, whether to reject the Null hypothesis?

2) Determine the fraction of the Null distribution that is more extreme than the observed difference: Complete up to 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.

Conventionally, if $p < 0.05$, we reject the Null hypothesis

(p=0.05 means 5% of Null distribution is more extreme than our observed outcome)

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.

Conventionally, if $p < 0.05$, we reject the Null hypothesis
(p=0.05 means 5% of Null distribution is more extreme than our observed outcome)

CAUTION!!!! The fact that $p \neq 0$ tells us that there is still a chance that we obtained our outcome by chance, so that the Null hypothesis is actually correct (and rejecting it is wrong).

Conventionally, if $p < 0.05$, we reject the Null hypothesis
(p=0.05 means 5% of Null distribution is more extreme than our observed outcome)

Convention leads to incorrect conclusions

Interpret p-values on 'sliding scale':

p ~ 1	Weaker evidence	Stronger evidence	p ~ 0

Conventionally, if $p < 0.05$, we reject the Null hypothesis
(p=0.05 means 5% of Null distribution is more extreme than our observed outcome)

Convention leads to incorrect conclusions

Interpret p-values on 'sliding scale':

p ~ 1	Weaker evidence	Stronger evidence	p ~ 0

p ~ 0.005 'substantial' or 'strong' evidence to reject Ho

p ~ 0.05 to 0.005 'suggestive' or 'moderate' evidence to reject Ho

p ~ 0.05 'suggestive' or 'moderate' evidence to reject Ho

p notable above 0.05 'weak evidence to reject Ho'

Given our data, what do we conclude?

Does average ‘Leaf date’ differ between Northern and Southern populations of Beech in the UK?

Differentiate between a Null and Alternative hypothesis.

Differentiate between a ‘sample’ and a ‘population’.

Consider how to sample appropriate data.

Appreciate the concept of ‘statistical significance’ and its pitfalls.

Explain how statistics uses a ‘Null Distribution’.

Conduct a Randomization Test to evaluate evidence whether two means differ.