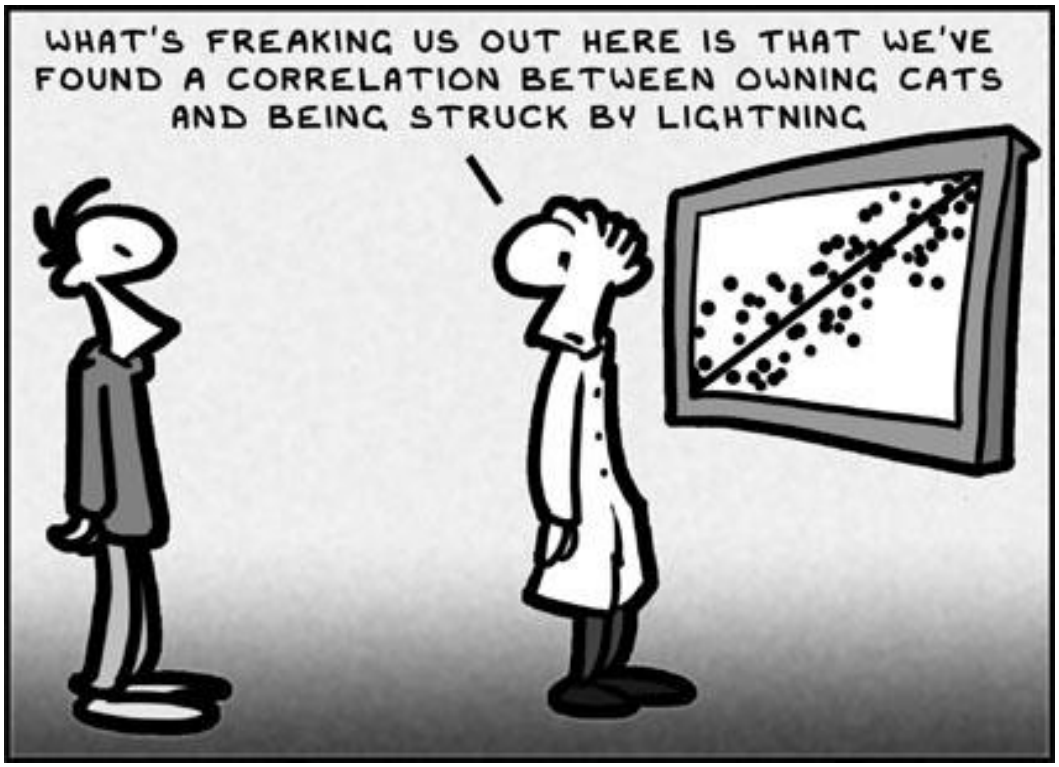



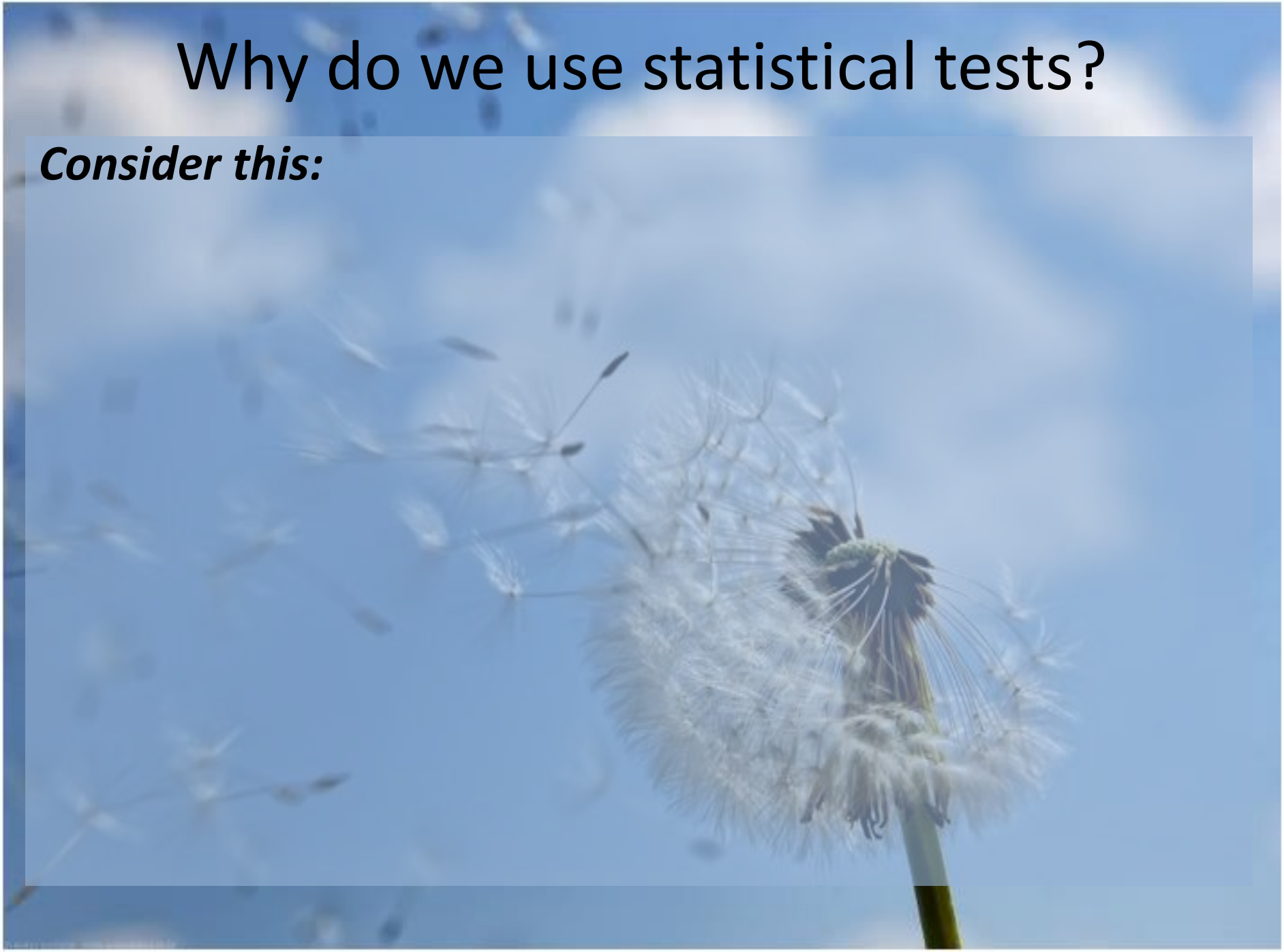

# Introduction to Statistics!

# Why do we use statistical tests?

# Why do we use statistical tests?

# Why do we use statistical tests?

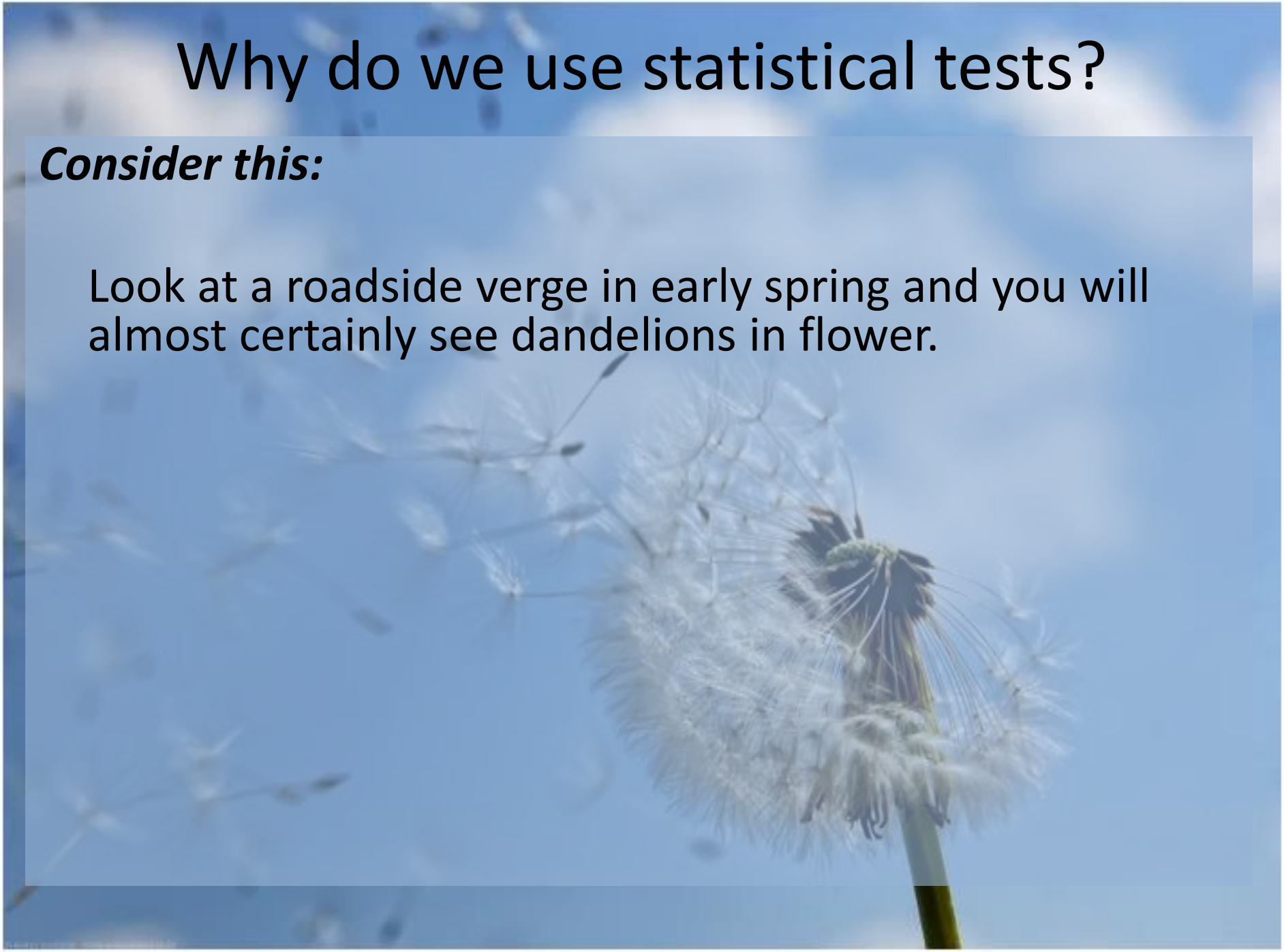
***Consider this:***



# Why do we use statistical tests?

***Consider this:***

Look at a roadside verge in early spring and you will almost certainly see dandelions in flower.

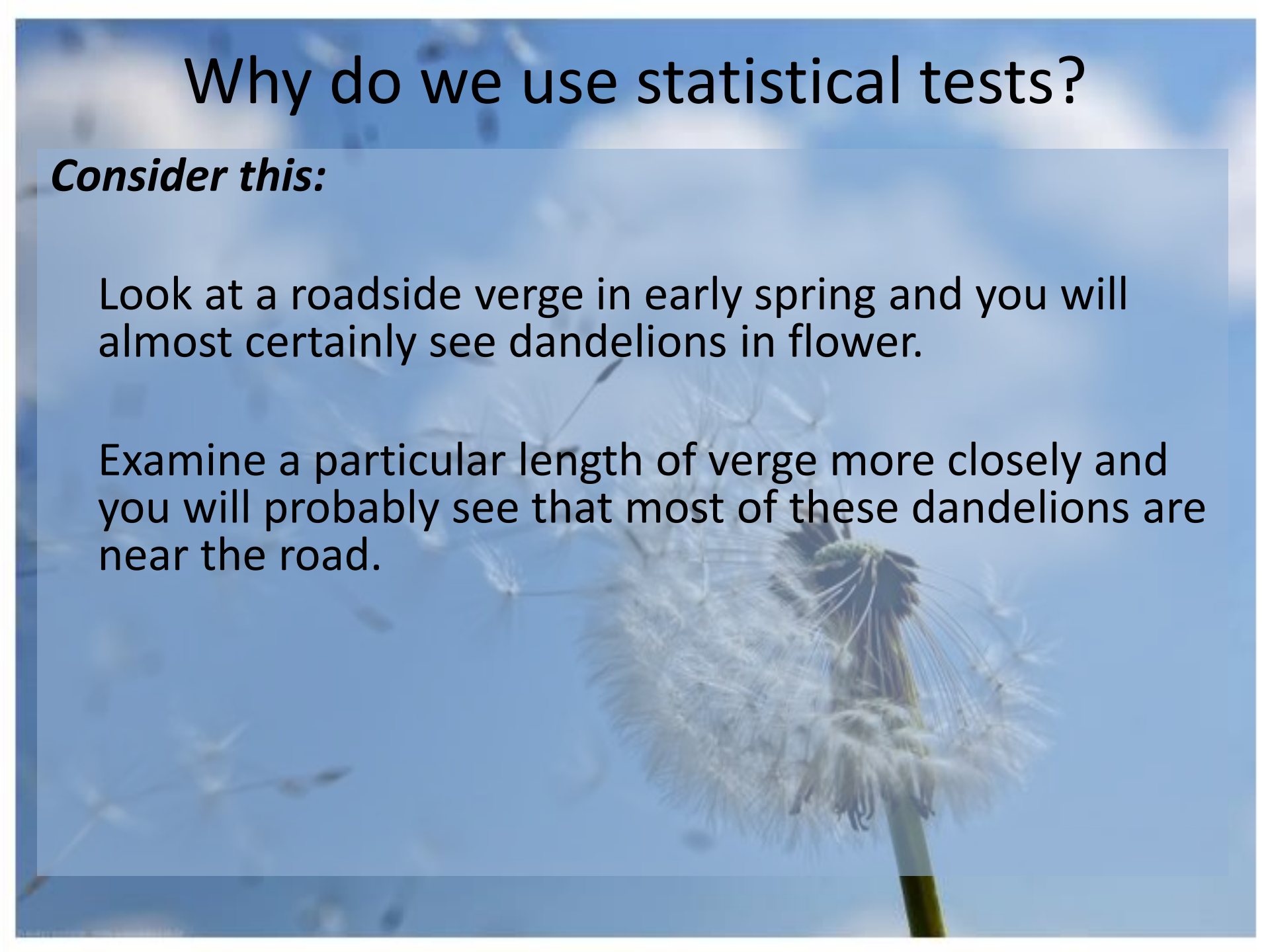


# Why do we use statistical tests?

***Consider this:***

Look at a roadside verge in early spring and you will almost certainly see dandelions in flower.

Examine a particular length of verge more closely and you will probably see that most of these dandelions are near the road.



# Why do we use statistical tests?

***Consider this:***

Look at a roadside verge in early spring and you will almost certainly see dandelions in flower.

Examine a particular length of verge more closely and you will probably see that most of these dandelions are near the road.

Farther away, dandelions are less common and other species of plant are more numerous.

# Why do we use statistical tests?

***Consider this:***

Look at a roadside verge in early spring and you will almost certainly see dandelions in flower.

Examine a particular length of verge more closely and you will probably see that most of these dandelions are near the road.

Farther away, dandelions are less common and other species of plant are more numerous.

***How can we explain this observation?***

# There are a number of possible biological explanations

# There are a number of possible biological explanations

- Perhaps the road was treated with salt during the winter and dandelions can tolerate high salt concentrations better than other plants.

# There are a number of possible biological explanations

- Perhaps the road was treated with salt during the winter and dandelions can tolerate high salt concentrations better than other plants.

- Perhaps they can tolerate frequent mowing or trampling.

# There are a number of possible biological explanations

- Perhaps the road was treated with salt during the winter and dandelions can tolerate high salt concentrations better than other plants.
- Perhaps they can tolerate frequent mowing or trampling.
- Perhaps the wind generated by moving vehicles aids seed dispersal.

# However...

# However...

- There is another possible explanation, which has very little to do with biology.

# However...

- There is another possible explanation, which has very little to do with biology.
- Maybe it's all due to **chance**.

# However...

- There is another possible explanation, which has very little to do with biology.
- Maybe it's all due to **chance**.
- It could be that we just happened to pick an area where there were more dandelions growing closer to the road.


# Null hypotheses

All of the statistical tests in the A2 Biology course are based on a null hypothesis.

Students should be able to state a null hypothesis clearly and concisely. It must be related to the investigation concerned.

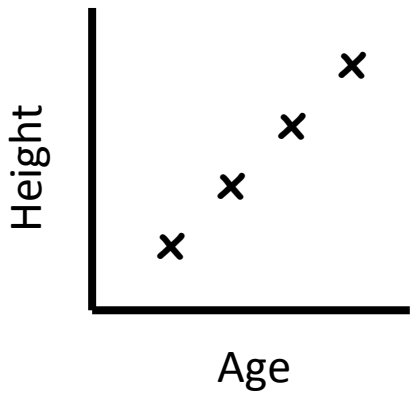
It is strongly recommended that it is worded in terms of:

**There is no significant difference between . . . and . .**
**There is no significant association between . . . and . .**
**There is no significant correlation between . . and . .**

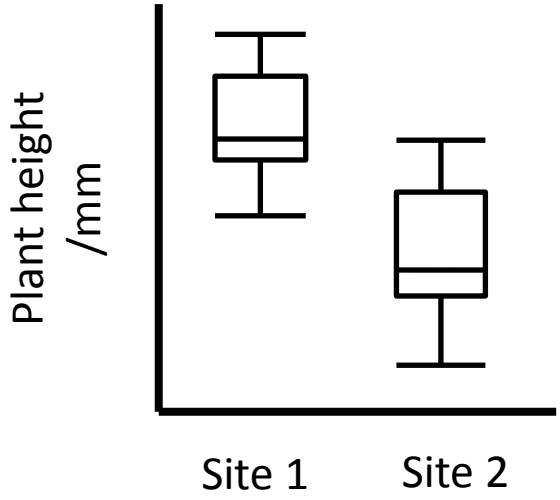
# Which statistical test do you use?

## What are you looking for?

### Relationship




### Difference



# Spearman’s Rank Coefficient of Correlation

BACK

# Spearman's Rank Coefficient of Correlation

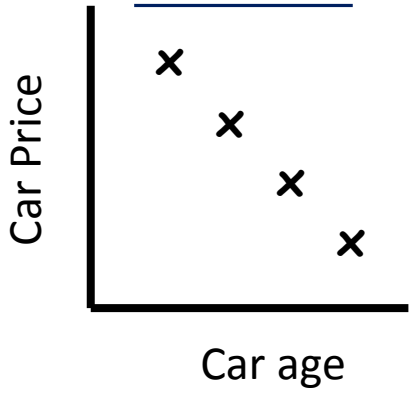
Looks for the strength of a relationship / correlation between two data sets

BACK

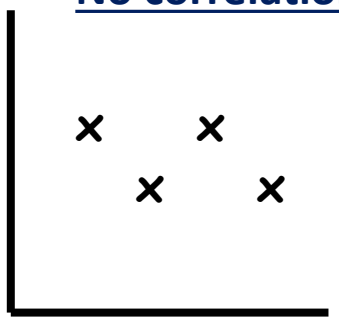
# Spearman's Rank Coefficient of Correlation

Looks for the strength of a relationship / correlation between two data sets

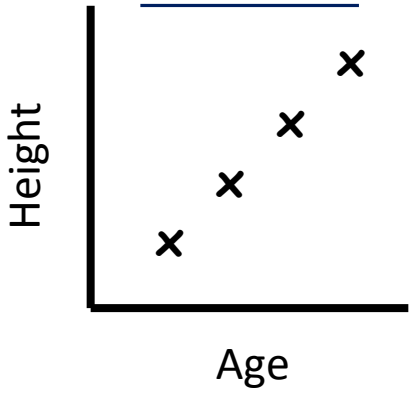
**Negative correlation**




**No correlation**



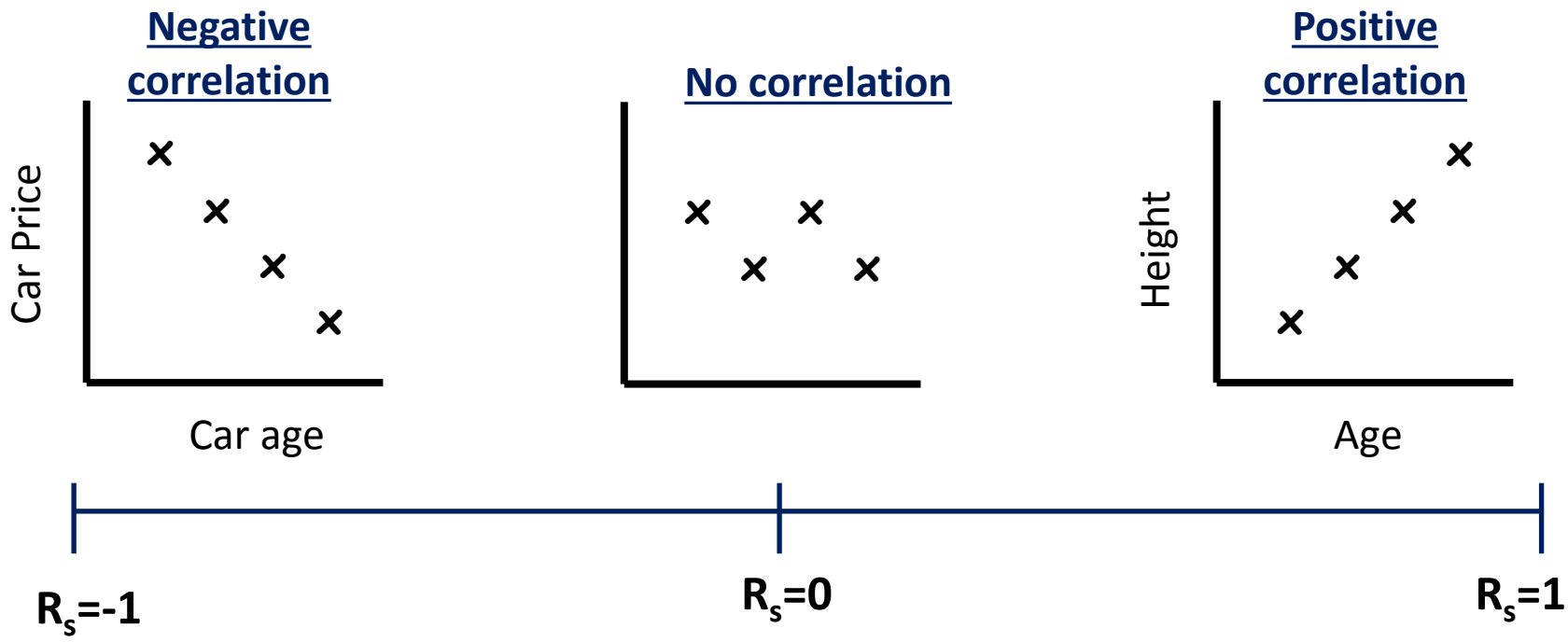
**Positive correlation**




BACK

# Spearman's Rank Coefficient of Correlation

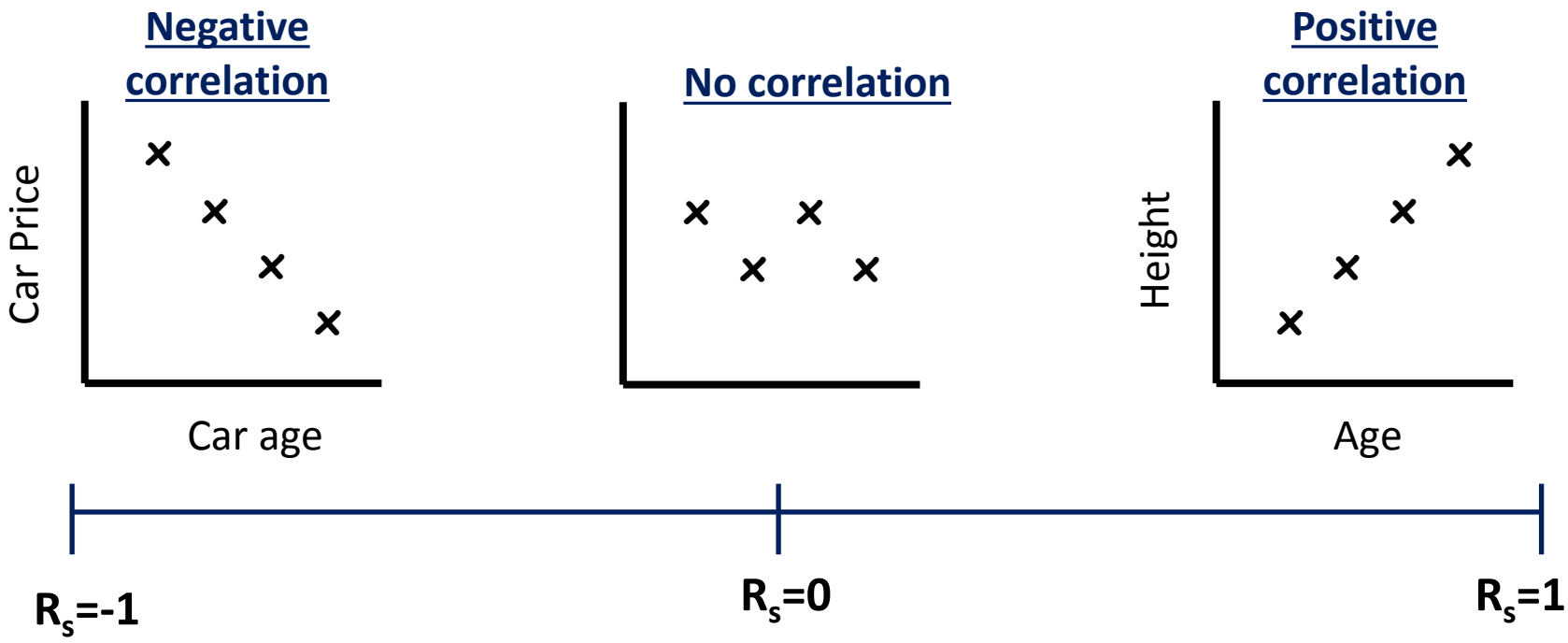
Looks for the strength of a relationship / correlation between two data sets




BACK

# Spearman's Rank Coefficient of Correlation

Looks for the strength of a relationship / correlation between two data sets



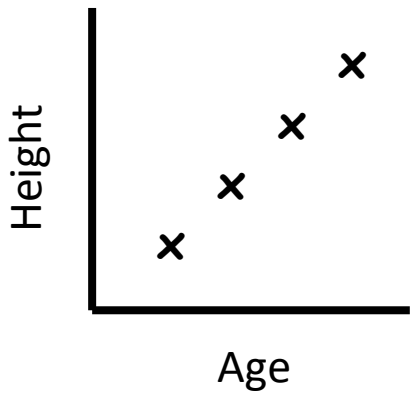

Requires a minimum of 10 sets of data (and a maximum of 30)

BACK

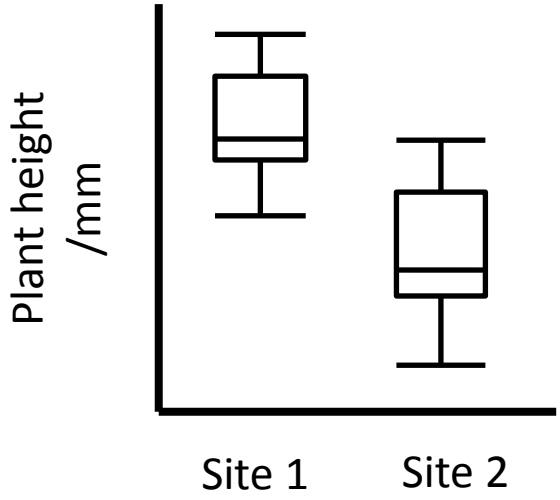
# Which statistical test do you use?

## What are you looking for?

### Relationship




### Difference



# Student’s t-test

# Student’s t-test

Looks for a difference between the means of two data sets

# Student's t-test

Looks for a difference between the means of two data sets

Data must be **normally distributed** and **continuous**

# Student's t-test

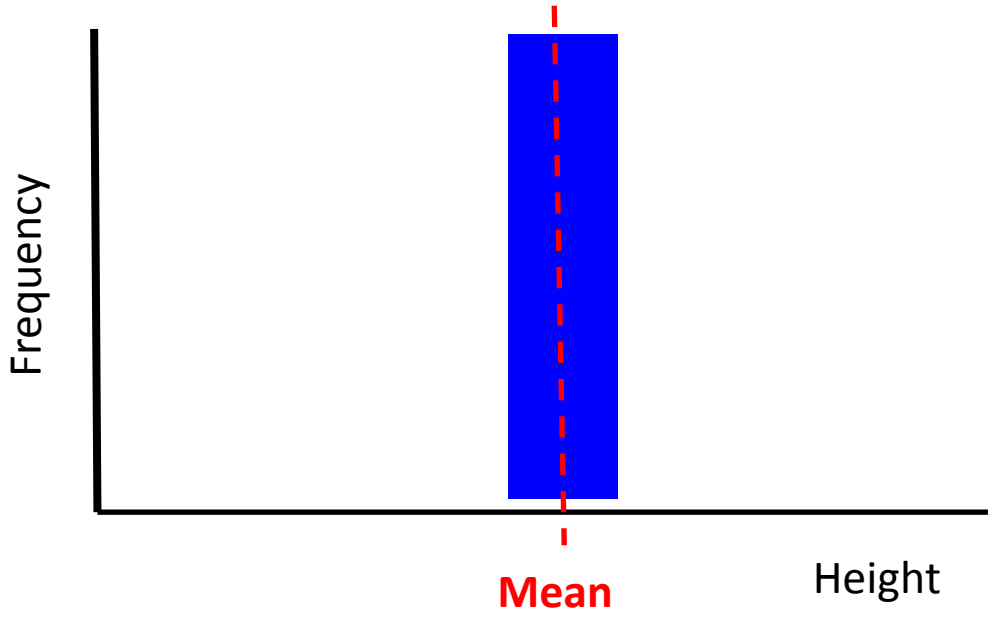
Looks for a difference between the means of two data sets

Data must be **normally distributed** and **continuous**

# Student's t-test

Looks for a difference between the means of two data sets

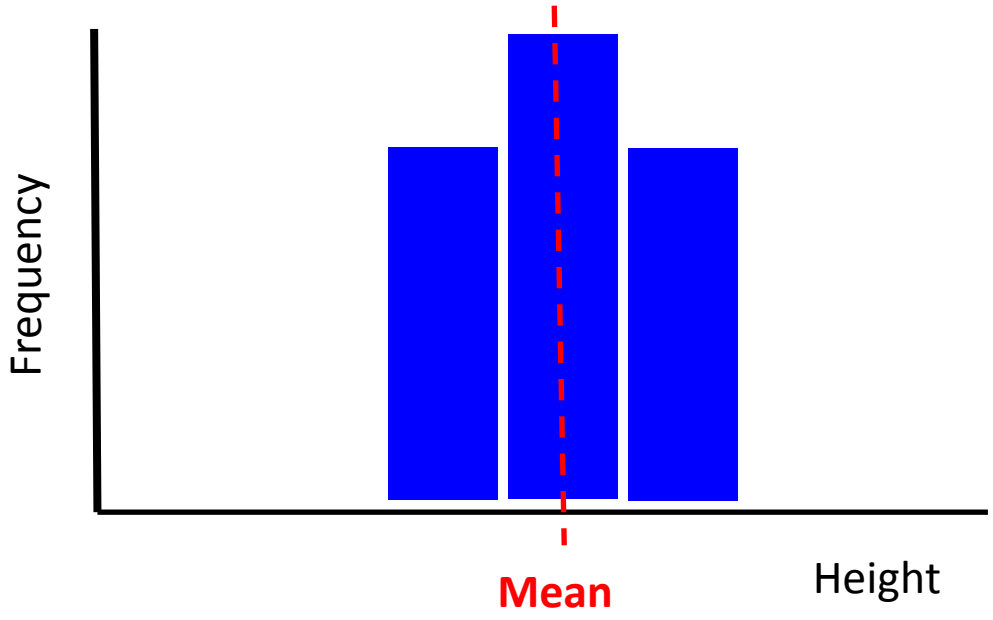
Data must be **normally distributed** and **continuous**



# Student's t-test

Looks for a difference between the means of two data sets

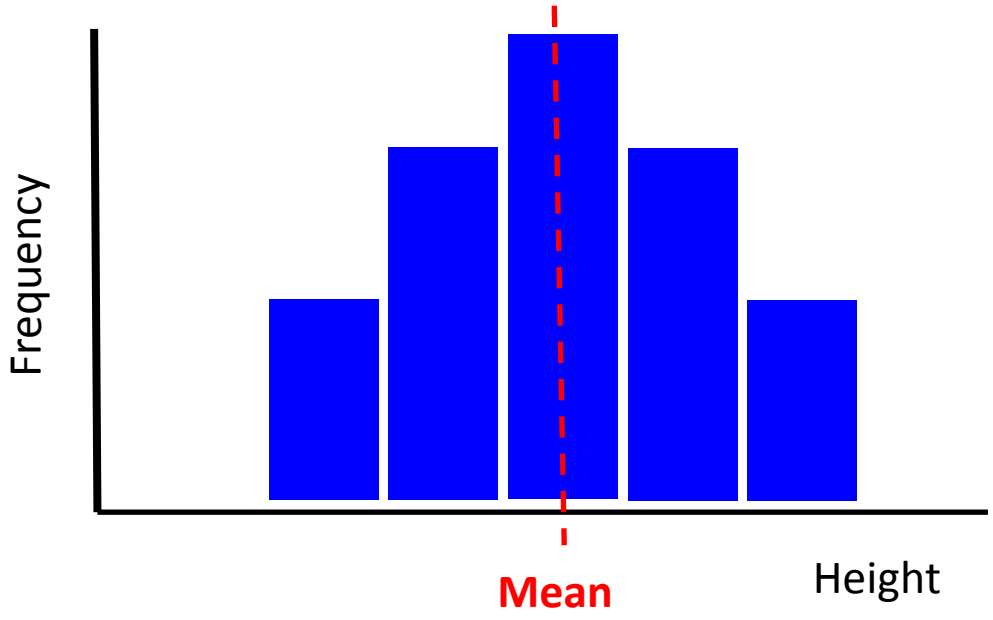
Data must be **normally distributed** and **continuous**



# Student's t-test

Looks for a difference between the means of two data sets

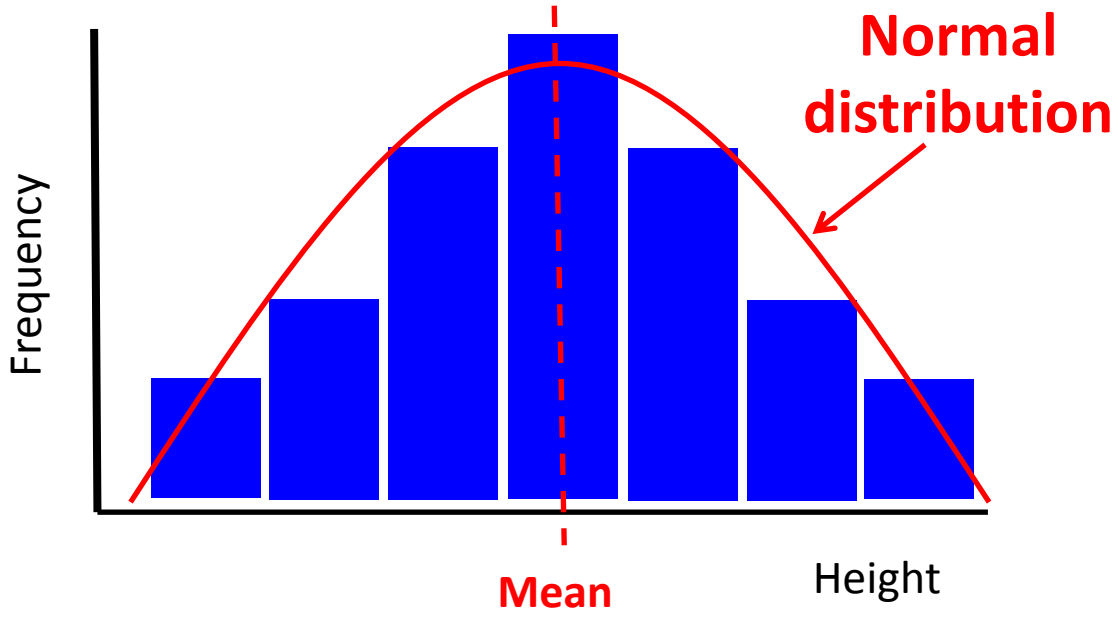
Data must be **normally distributed** and **continuous**



# Student's t-test

Looks for a difference between the means of two data sets

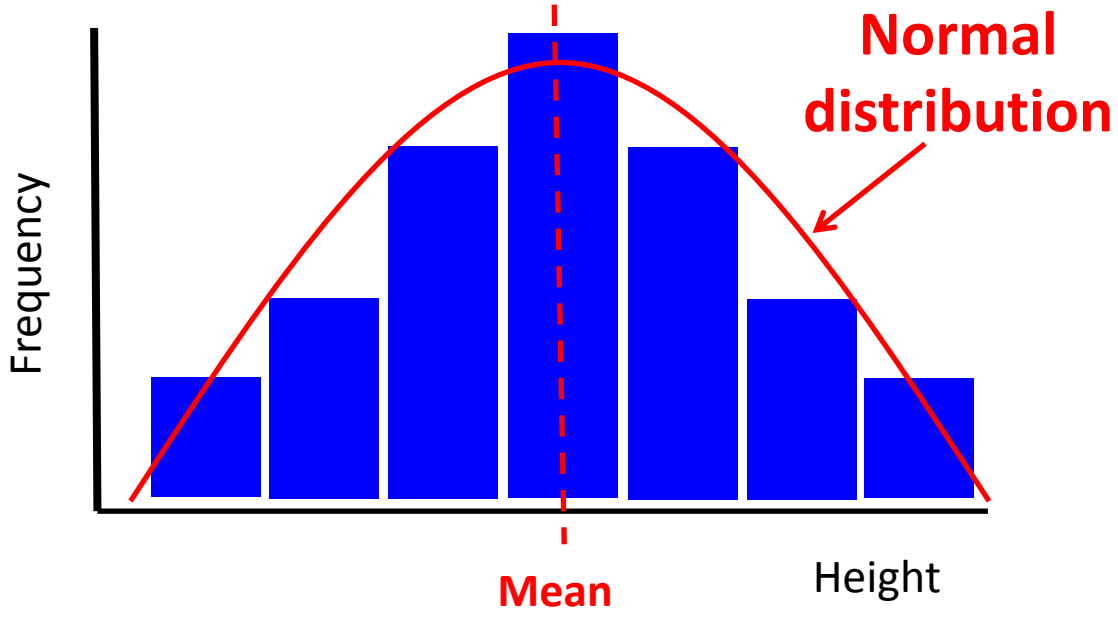
Data must be **normally distributed** and **continuous**

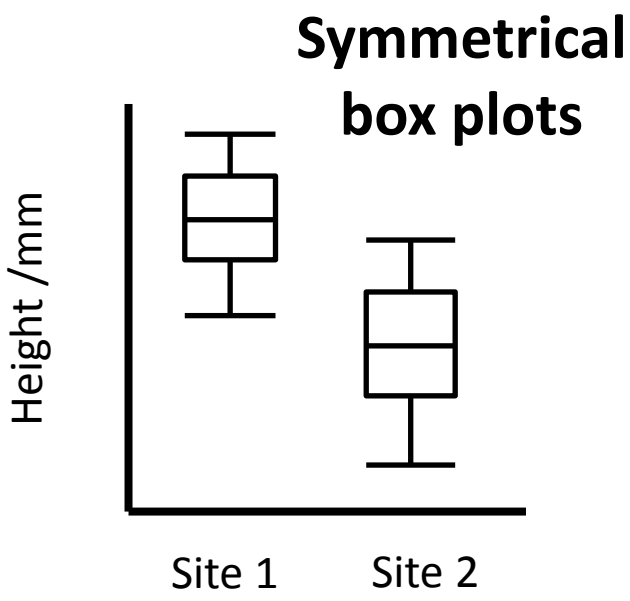


# Student's t-test

Looks for a difference between the means of two data sets

Data must be **normally distributed** and **continuous**

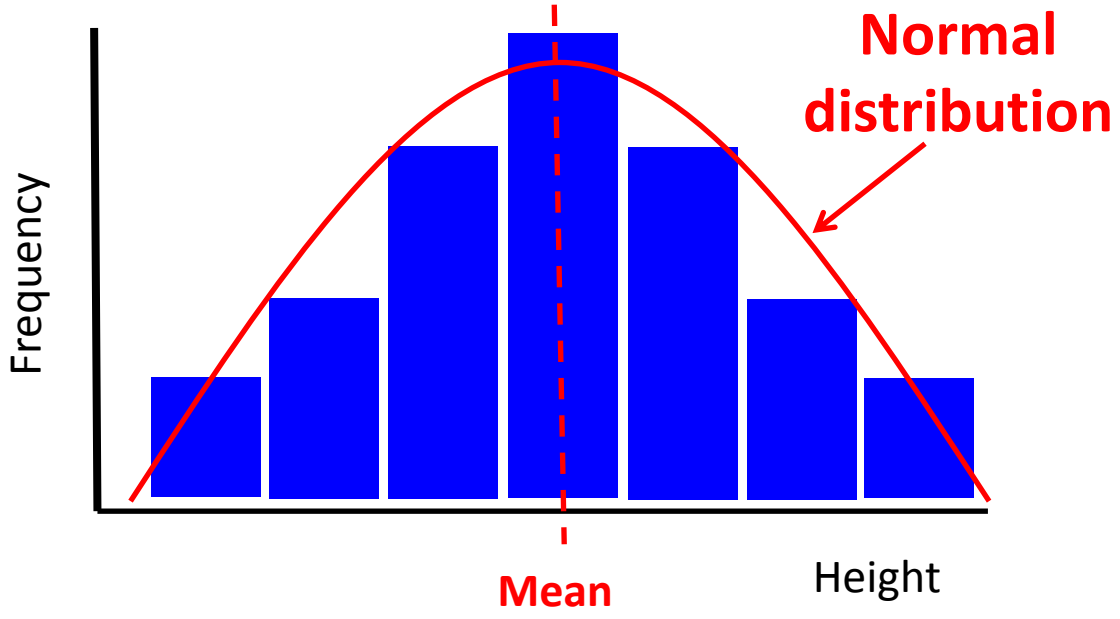


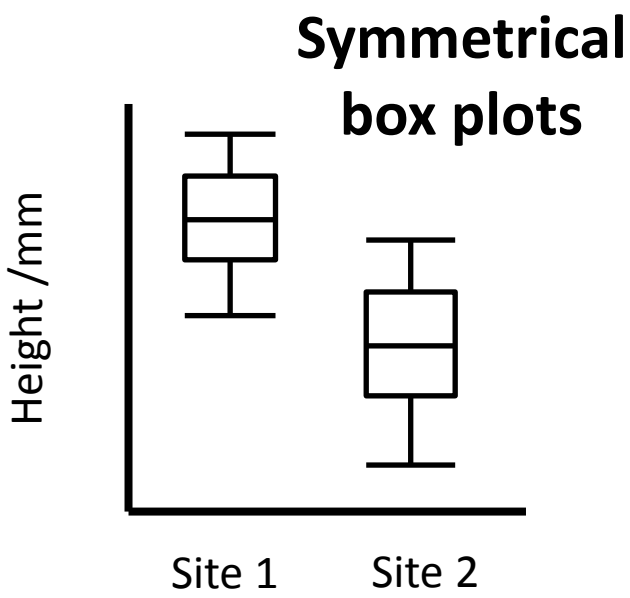



# Student's t-test

Looks for a difference between the means of two data sets

Data must be **normally distributed** and **continuous**




You need a minimum of 10 repeats in each data set and a maximum of 30

# Mann Whitney U Test

BACK

# Mann Whitney U Test

Looks for a difference between the means of two data sets

BACK

# Mann Whitney U Test

Looks for a difference between the means of two data sets

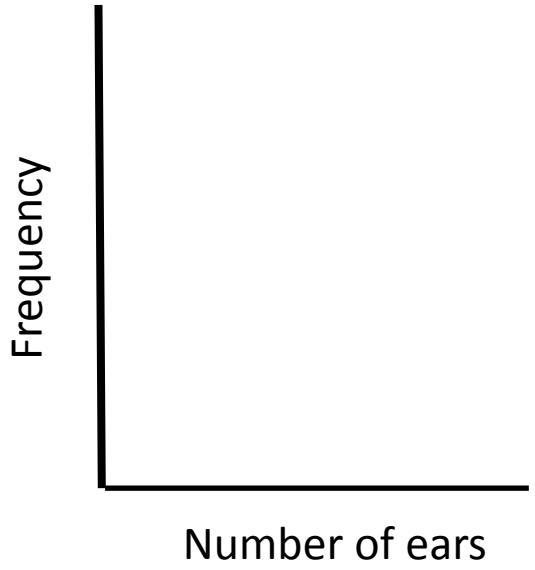
Used for data which is **not normally distributed (skewed)** and/or **not continuous (categorical)** e.g. shoe size

BACK

# Mann Whitney U Test

Looks for a difference between the means of two data sets

Used for data which is **not normally distributed (skewed)** and/or **not continuous (categorical)** e.g. shoe size

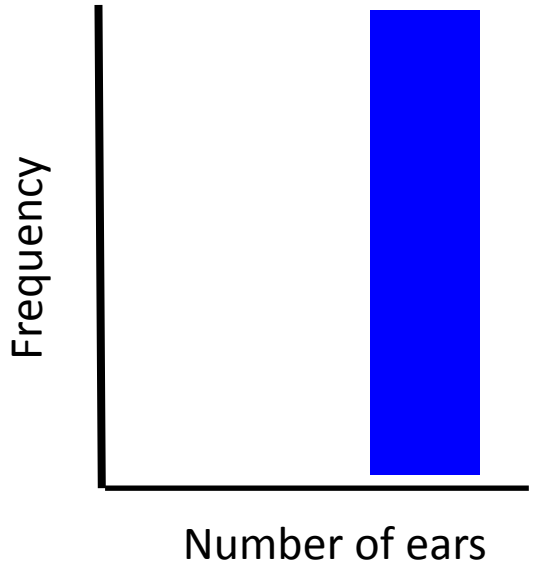



BACK

# Mann Whitney U Test

Looks for a difference between the means of two data sets

Used for data which is **not normally distributed (skewed)** and/or **not continuous (categorical)** e.g. shoe size

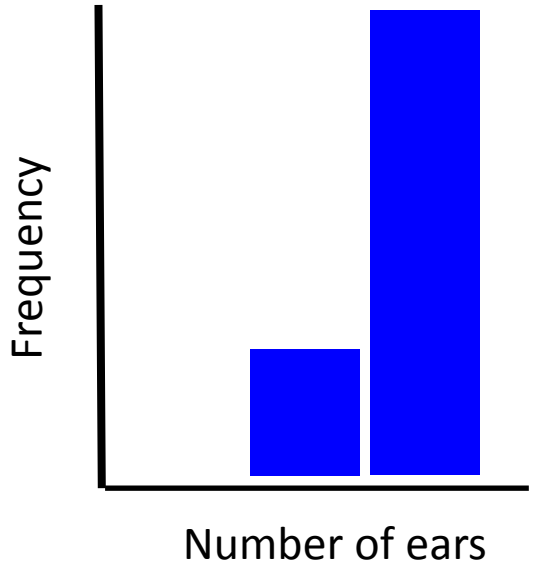



BACK

# Mann Whitney U Test

Looks for a difference between the means of two data sets

Used for data which is **not normally distributed (skewed)** and/or **not continuous (categorical)** e.g. shoe size

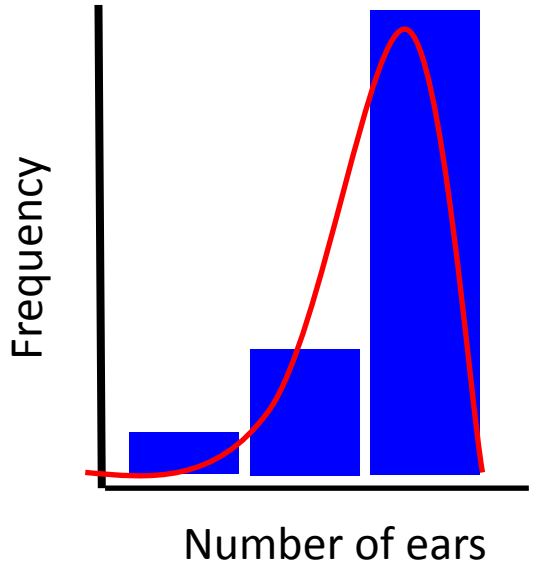



BACK

# Mann Whitney U Test

Looks for a difference between the means of two data sets

Used for data which is **not normally distributed (skewed)** and/or **not continuous (categorical)** e.g. shoe size

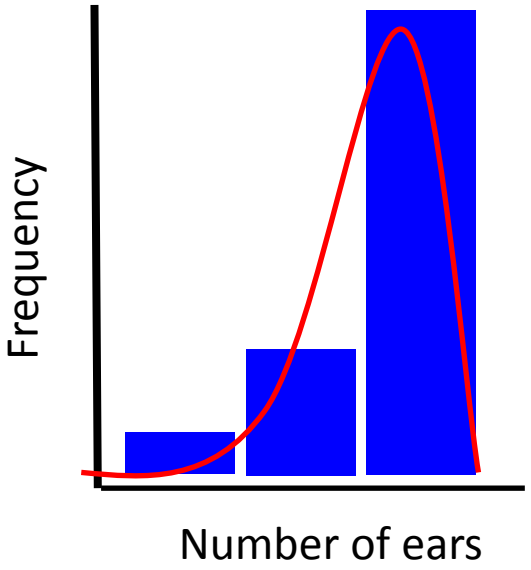


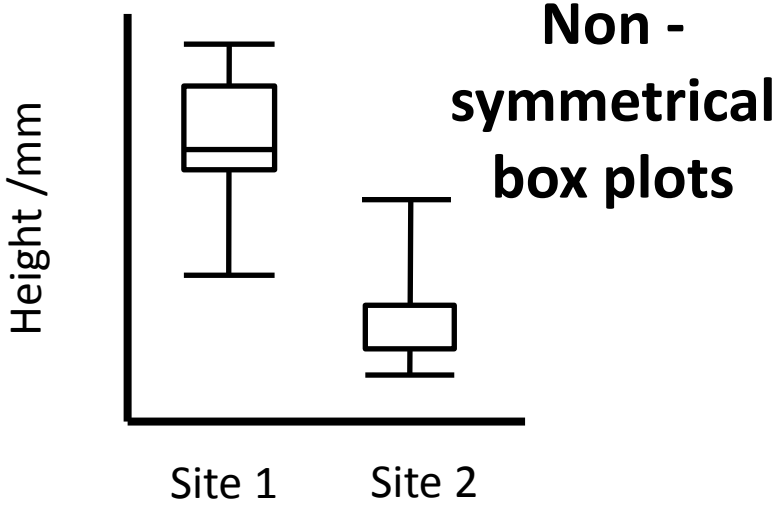

BACK

# Mann Whitney U Test

Looks for a difference between the means of two data sets

Used for data which is **not normally distributed (skewed)** and/or **not continuous (categorical)** e.g. shoe size



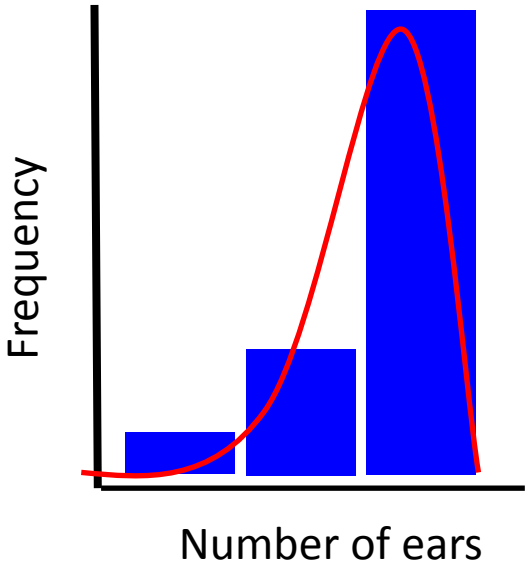

**Non - symmetrical box plots**

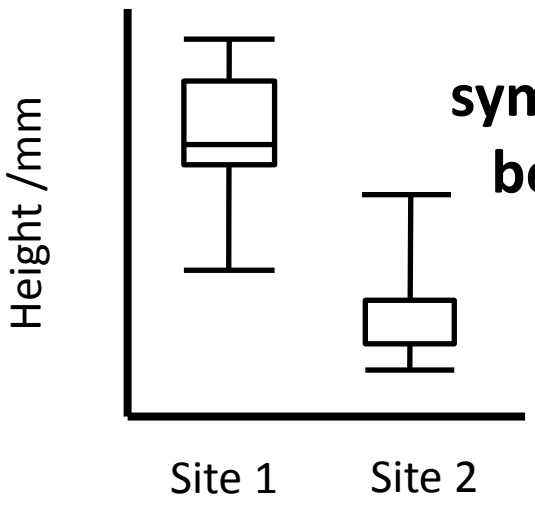
BACK

# Mann Whitney U Test

Looks for a difference between the means of two data sets

Used for data which is **not normally distributed (skewed)** and/or **not continuous (categorical)** e.g. shoe size



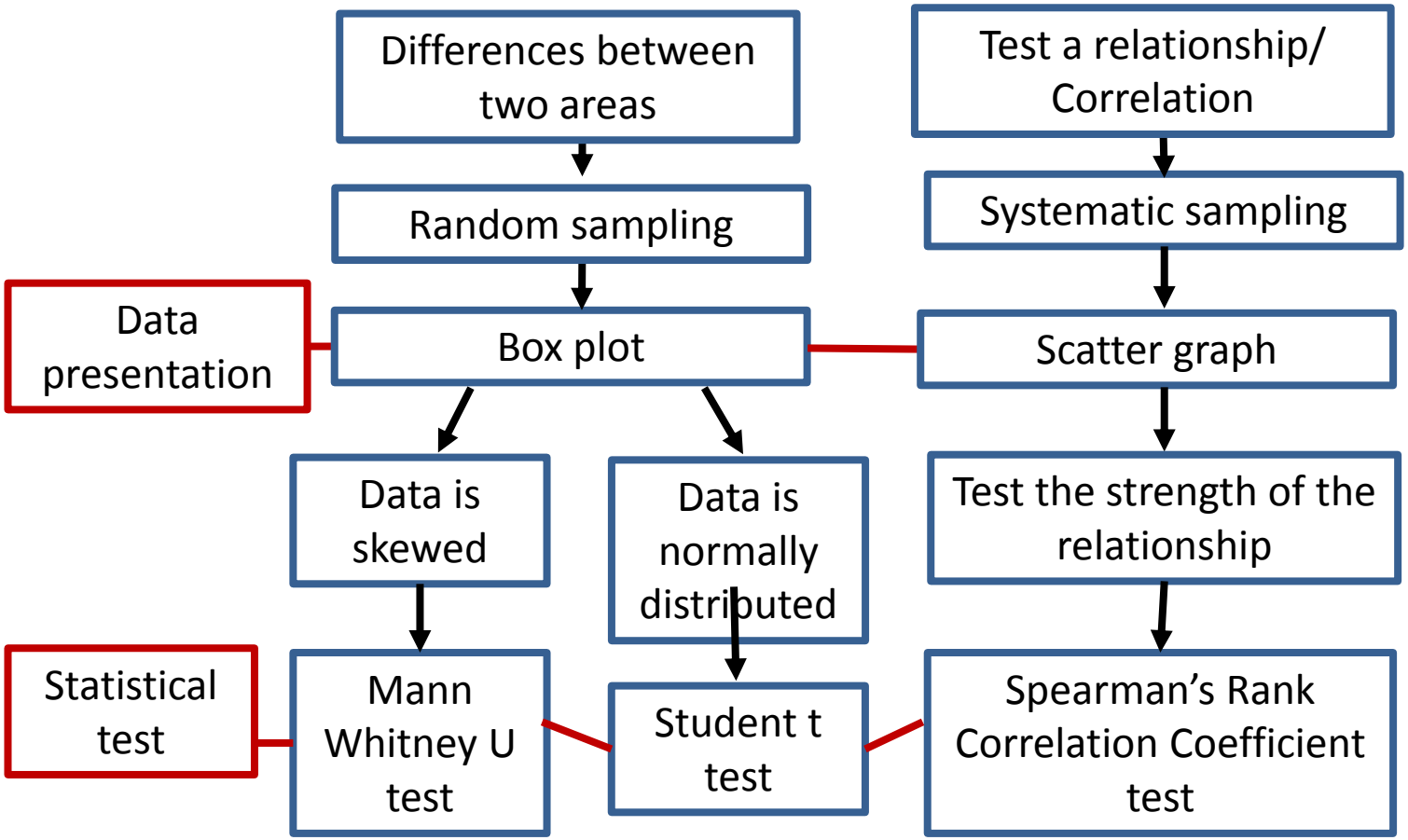

**Non - symmetrical box plots**

You need a minimum of 10 repeats in each data set and a maximum of 30

BACK

# Statistical tests

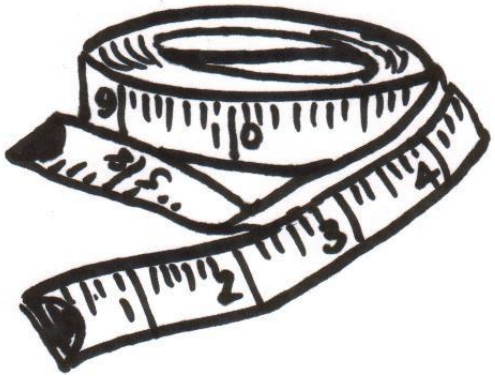
- Minimum sample size is at **least** 10 samples for a statistical test.
- How confident you are of the difference or relationship and whether it could just be down to chance



# Which statistical test do you use?

What sort of data will you obtain from your investigation?

## Measurements



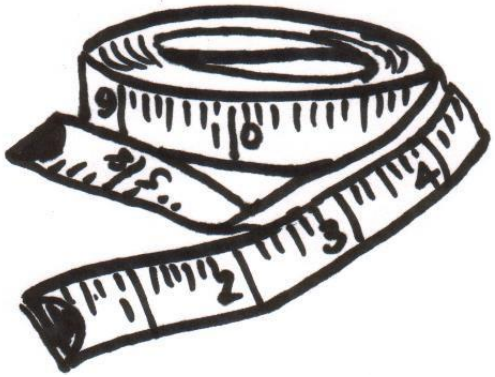
## Frequencies

| | Sparrows | Robins |
|---|---|---|
| Friday | I | 𝍸 |
| Thursday | 𝍸 I | 𝍸 |
| Wednesday | II | I |
| Tuesday | 𝍸 | II |
| Monday | II | I |
| |  |  |

# Which statistical test do you use?

What sort of data will you obtain from your investigation?

## Measurements



## Frequencies

| | Sparrows | Robins |
|---|---|---|
| Friday | I | 𝍸 |
| Thursday | 𝍸 I | 𝍸 |
| Wednesday | II | I |
| Tuesday | 𝍸 | II |
| Monday | II | I |
| |  |  |

# Chi-squared ($\chi^2$) Test

BACK

# Chi-squared ($\chi^2$) Test

Looks for the strength of an association between different **categories**

BACK

# Chi-squared ($\chi^2$) Test

Looks for the strength of an association between different **categories**

| |  |  |
|---|---|---|
|  | **90** | **10** |
|  | **15** | **75** |

BACK

# Chi-squared ($\chi^2$) Test

Looks for the strength of an association between different **categories**

| |  |  |
|---|---|---|
|  | **90** | **10** |
|  | **15** | **75** |

Requires a minimum of 50 repeats

BACK