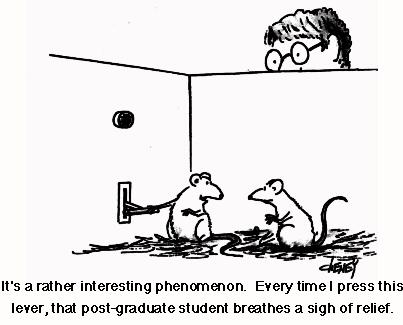
Level 3 Experim**e**nts



# By Liz Sneddon

# Name:\_\_\_\_\_\_\_\_\_\_\_\_\_Vocabulary List

|  |  |  |
| --- | --- | --- |
| **Bias** |  | Something that causes favouritism. |
| **Blinding** |  | Where the experimental unit does not know which treatment they have received.  Double blinding is when both the experimental unit and the doctor don’t know which treatment is being received. |
| **Cause** |  | This is usually the *treatment*. |
| **Context** |  | The real world story or facts behind an experiment. |
| **Control group** |  | The group who does not receive the treatment. |
| **Effect** |  | The outcome of applying a treatment, measured by the *response variable*. |
| **Experiment** |  | Process of planning, running, and looking at the results of a test. |
| **Experimental Group**  **or**  **Treatment Group** |  | Group of experimental units.  or  The group who receives the treatment |
| **Experimental Unit** |  | Single person who is being tested upon in an experiment. |
| **Experimenter** |  | Person or group of people in charge of running an experiment. |
| **Hypothesis** |  | Predication, or expectation. Usually made before an experiment. |
| **Independent variable** |  | Usually takes only two values, *placebo* and *treatment.* |
| **Median** |  | The central or middle value of an ordered dataset |
| **Placebo** |  | Simply put, a fake *treatment.* |
| **Purpose** |  | The thoroughly developed line of reasons for running an experiment. |
| **Random Allocation** |  | Process of randomly assigning *experimental units* to groups using, for example a deck of cards or flipping a coin. |
| **Randomisation test** |  | Process of testing if chance alone is influencing the results from an experiment. |
| **Response variable**  **Or**  **Dependent variable** |  | The measurement that is the main focus of an experiment. |
| **Spread** |  | The spread of the data around the median, measured by the interquartile range (IQR) or standard deviation. |
| **Treatment** |  | An applied change or influence that should result in a change in the *response variable.* |
| **Variable** |  | A measurement, or characteristic (e.g. weight or gender). |

## What is an experiment?

Watch the ghostbusters video, and answer the following questions:

1. What was the experiment testing for?

2. What would Dr. Venkman have been recording (writing down)?

3. What was the treatment or stimulus?

4. Is it possible for the subject to have just guessed the correct answer?

5. The experiment is repeated 80 times for each person. If there are 5 possible answers, how many out of 80 would you expect them to get right if they just guessed?

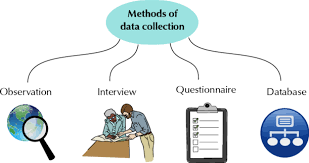
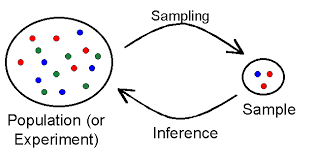
6. How many out of 80 would they need to get right to show evidence of Extra Sensory Perception?

7. Is it a well-designed experiment? Why/why not?

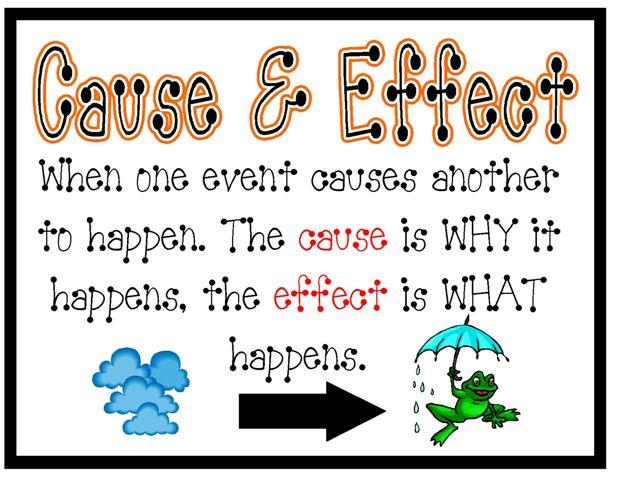
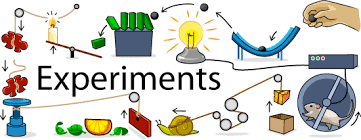
## Observational study versus Experimental study

Watch the video explaining the difference between an observational study and an experimental study.

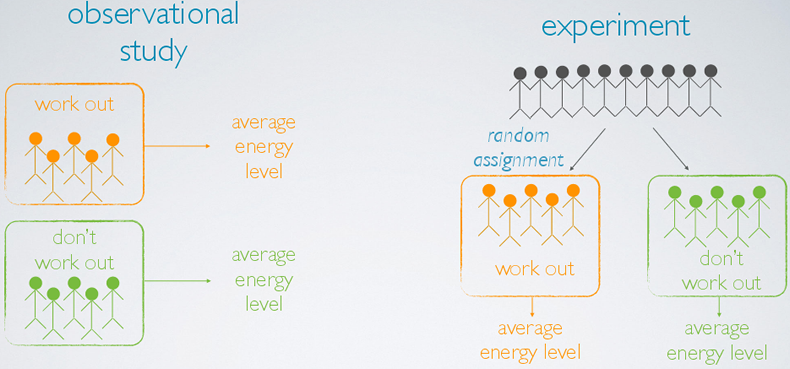
With an **observational study** we can make **inferences:**

With an **experiment** we can show **cause and effect**.



The difference between these is due to the **random allocation**:



## Experiments

In this topic we are only concerned with true experiments. What classifies an investigation as an experiment? There must be an intervention. The experimenter must change **just one thing** between the groups being studied. All other conditions must be controlled.

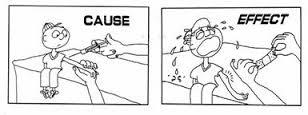
**Exercise:**

What is an observational study?

What is an experiment?

# Problem

In your investigation question you need the following parts (and be quite specific):

* Groups
* Measure
* Population described
* The word “**effect**”

You also need to have a purpose and make a prediction about what you think will happen.

### Example

If we are interested in doing an experiment to see if senior students can stand on their dominant leg or non-dominant leg for longer, our investigation question might be:

I wonder if standing on your dominant leg has an **effect** on the length of time (seconds) you can stand on this leg, for students in STA3 at Ormiston Senior College, in 201..

Our purpose might be:

As a PE teacher, I want to know whether or not students are stronger on the dominant side of their body. If they are, I have a strength training exercise to develop the non-dominant side of their body, so that both sides of their body are equal in strength. To test this, I will make students stand on their dominant or non-dominant leg, and see if there is a difference in how long they can stand.

Our prediction might be:

I think students can stand on their dominant leg for longer because the muscles on their dominant side of their body are used more, and are stronger, therefore they will stand for longer on their dominant leg.

Some research to support this:

“"It is extremely common for people to have strength differences between their sides," says Chris Powell, CSCS, celebrity trainer and DietBet.com game host. "In fact, it is more uncommon for our bodies to be perfectly symmetrical in size and strength than it is for them to be different."

We unconsciously use our dominant side far more than our weak side. This can be pushing or pulling doors open, rolling over to push yourself out of bed, or the side you always chose to take the first step onto the stairs," says Powell. "While we wouldn't necessarily consider this every activity 'exercise,' the more we repeatedly use one side, the more efficiently our brain learns to fire to those particular muscles. This results in stronger muscles on that side, and quite often larger muscles as well.”

These quotes support the idea that students might be stand for longer on their dominant leg, and the quotes were taken from: <http://www.fitnessmagazine.com/weight-loss/tips/advice/why-your-right-side-is-stronger-than-your-left/>

## Experiment 1

We want to do an experiment with the class, to find out if there is any difference in how long students can write a paragraph with their dominant hand. We want to compare students who write with their dominant hand with those who write with their non-dominant hand, and see if there is any difference between how long it takes them to write the same paragraph.

Write an investigation question, purpose and a prediction for this situation.

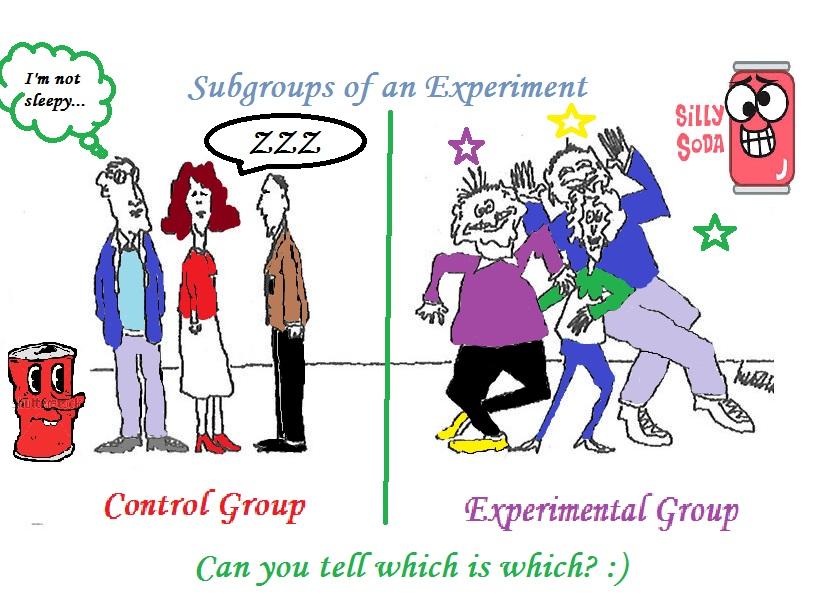
# Plan

Your plan needs to include:

* Stating your experimental units (who is going to do the experiment)
* Describing the treatment group and control group.
* Describing the response variable (what you are measuring, and its units).
* Describing the random allocation of treatment to the experimental units (how are you going to randomly allocate students to be in either the treatment or control group?)
* Describe the data collection and recording methods (what is the data you will collect, how will you collect it, where and how will you record this information?)
* Describe any possible sources of variation (what are some factors that you need to control, so that the only difference in the treatment is the **ONE** factor you are investigating).

## Control Group

Experiments usually have a **control group**, a group that receives **no treatment** or receives an **existing or established treatment**. This allows any differences in the response, on average, between the control group and the other group(s) to be visible.

For example, we may compare a new cancer treatment with the current cancer treatment (control group).

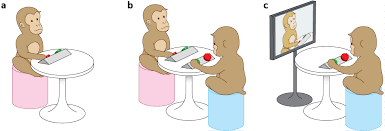
Or we may compare a new medication with a placebo (sugar pill – control group).

When the groups are similar in all ways apart from the treatment received, then any observed differences in the response (if large enough) among the groups, on average, is said to be caused by the treatment.

## Experimental design principles

Issues that need to be considered when planning an *experiment.*

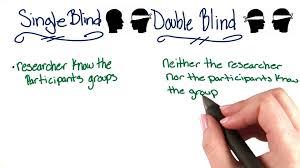
The following issues are the most important:



**Controlled conditions:**

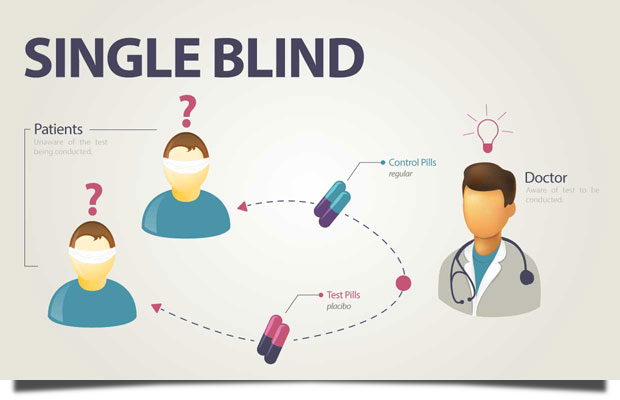
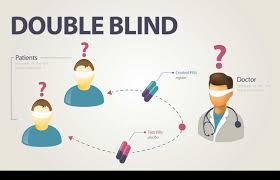
Consider carefully whether you have maintained controlled conditions for each treatment.

**Blinding and double blinding:**

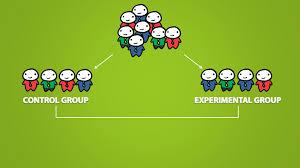
Were any individuals able to observe someone else do the experiment before they did it? Ideally, no subject should know anything about the experiment before they do it, as the knowledge might affect how they respond to the treatment.

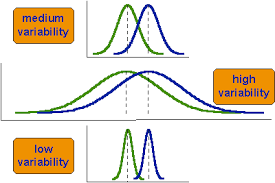
There are two types of blinding:

* single blind and
* double blind.

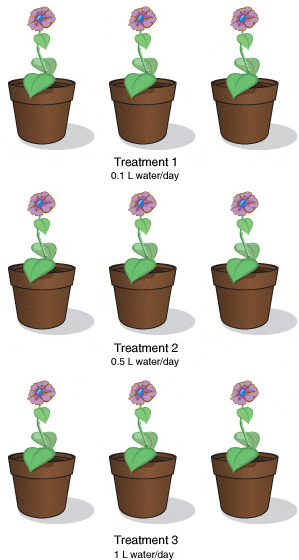
 

**Randomisation**:

A randomising method should be used to allocate individuals to groups to try to ensure that all groups are similar in all characteristics apart from the treatment received. The larger the group sizes, the better the balancing of the characteristics, through randomisation, is likely to be.

**Variability**: 

A well-designed experiment attempts to minimise unnecessary variability. The use of random allocation of individuals to groups reduces variability, as does larger group sizes. Keeping experimental conditions as constant as possible also restricts variability.

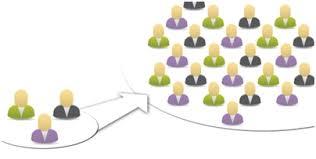


**Replication:**

For some experiments, it may be appropriate to carry out repeated measurements. Taking repeated measurements of the *response variable* for each selected value of the *explanatory variable* is good experimental practice because it provides insight into the variability of the response variable.

## Random allocation

When we are comparing 2 groups of people, we need to be sure that both groups are very similar in everything except the one variable we want to measure.



For example, we want all the following factors to be similar in both groups:

• Ethnicity (in case one ethnic group have stronger leg muscles than other ethnic groups)

• Gender (as males may have stronger leg muscles than females)

• Time of day that the test is done (people might be able to stand for longer on one leg in the morning compared to the evening)

• Environment (the same area, so it is flat and fair)

• Students spaced equally apart so they don’t fall and knock each other over.

• Etc.

It isn’t easy to make sure all these and other factors are similar in both groups, but there is another way we can do this. By randomly allocating people into the two groups, this will randomly balance these other factors.

**How do we do this?**

The easiest way is to have pieces of paper with either A or B written on it, and allocate A to be one group and B to be the second group.

You could also use other methods, such as:

* Use a deck of cards – black cards versus red cards
* Tossing a coin,
* Rolling a die,
* Generating a random number on the calculator,
* Etc.

### Example

1. I have 30 students that I want to put into 2 test groups. I select 15 black cards from a deck of cards. These will represent group 1.
2. I then select 15 red cards, these will represent group 2.
3. I shuffle the cards so that they are randomly mixed.
4. I go up to student 1 and give them a card. If it is black they will go into group 1, and if it is red they will go into group 2.
5. I hand out a card to each student and then move them into the 2 groups.

## Sample size

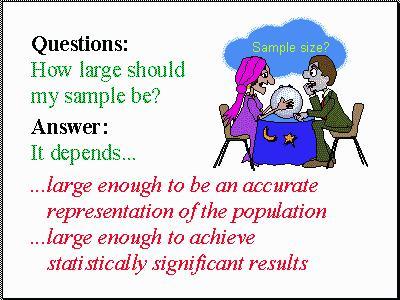
You have to decide on your sample size.

For **count** data: use a sample size of **50**.

For **measurement** data: use a sample size of **30**.

**Remember**: you will need these sample sizes in **each** group.

**Note**: while we would like a sample size of 50 for count data, the constraints of us doing an investigation using 50 people in each group means we have to compromise.



## Recording data

The team needs to decide on how to set up your data table. You need to think about what data you are going to write down, and what columns you need to have (including headers).

You will want one column with the treatment, and one with the response variable.

|  |  |
| --- | --- |
| **Treatment variable** | **Response Variable (units)** |
| Treatment group / Control group |  |
| … |  |

### Example

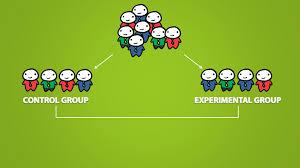
If our investigation question is:

I wonder if standing on your dominant leg has an **effect** on the length of time (seconds) you can stand on this leg, for students in STA3 at Ormiston Senior College, in 2017.

**Experimental Units:**

Students in 13STA at McAuley High School, in 2016.

**Treatment variable:**

The treatment group is: the group of students (as described above) who will be standing on their non-dominant leg for the experiment.

The control group is: the group of students (as described above) who will be standing on their dominant leg for the experiment.

**Response variable:**

I will be measuring the length of time that students (as described above) are able to stand on either their dominant or non-dominant leg. I will measure this with a stopwatch, and measure it in the number of seconds.

I would like to get students to repeat this experiment twice.

**Random allocation:**

I will be randomly allocating the students either into the treatment group (standing on their non-dominant leg), or the control group (standing on their dominant leg).

I will do this by finding the sample size of students present, dividing this in half, and then writing group A or B on this number of separate pieces of paper. For example, if 30 students are present, then I would have 15 pieces of paper with A written on it, and 15 pieces of paper with B written on it.

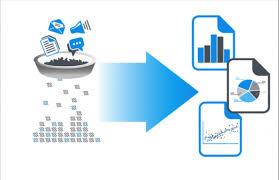
I then need to randomly shuffle these pieces of paper before handing them out to the students.

I will not tell the students whether they are in the treatment group or control group. Instead I will blind the students, and test group A first (while group B leave the room), and then test group B second (while group A leave the room).

For the purposes of this experiment, group A will be the treatment group, and these students will be standing on their non-dominant leg. While group B will be the control group, and these students will be standing on their dominant leg.

I will be randomly allocating students into these two groups, so that all other characteristics (such as height, weight, age, ethnicity, etc) are balanced, and there is no bias present in the allocation of students into the groups. This will ensure that I have a properly designed experiment and not an observational study. This means that both the treatment group and control group should be similar in all characteristics, and therefore the only difference in the results (the time standing on one leg) should be due to the treatment (whether it is the dominant or non-dominant leg) and not any other factors or characteristics.

**Data collection:**

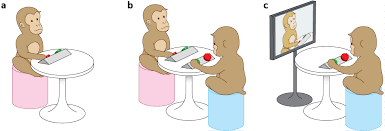
A large stopwatch will be displayed on the whiteboard by the projector. Students will be given instructions to record their time onto their pieces of paper with A or B written on it. At the end of the experiment, students will hand these pieces of paper into me, and I will record the data on a Google Sheet, as follows:

|  |  |
| --- | --- |
| **Group** | **Time standing on one leg (seconds)** |
| Dominant leg / Non-dominant leg |  |
| … |  |

**Sources of variation (controlled factors):**

Some of the factors I will control are:

* Giving the instructions clearly at the start, so that all students are clear that they need to track the stop watch time precisely. Otherwise some students may be busy talking to each other and not notice their time for a few seconds after they drop their leg.
* I will make sure that all students are standing at least an arms length away from each other, so that if one student loses their balance, this won’t knock the student next to them over.
* I will demonstrate to the students at the beginning how they should place their foot against the side of their calf muscle, so that all students are holding their legs in the same position, so the amount of muscle fatigue should be similar.



## Experiment 1

Write out a plan for your experiment.

# Data

## Collecting data

For the assessment you will need to work in groups of around 3-4 students on an investigation. This means that you will work together to run your experiment with a group of students.

The different roles that need to be allocated are:

* Presenter (talking to the students and giving instructions)
* Data Recorder (recording the students data into a table)
* Materials manager (making sure that the group has all the equipment it needs)
* Observation recorder (writing down any observations while the students are doing the investigation)

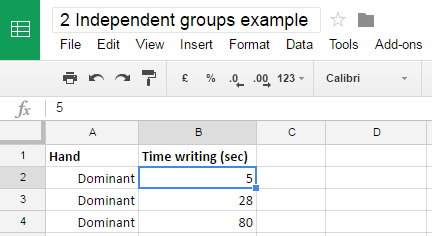
### Entering data into Excel

We will go on to use NZGrapher to analyse our data. This means that we need the data entered into Google Sheets and then copied into NZGrapher (using the “**Paste Table**” button).

Alternatively, you can enter the data directly into NZGrapher by choosing the “Empty dataset for editing” under the Data Source (top right of the screen).

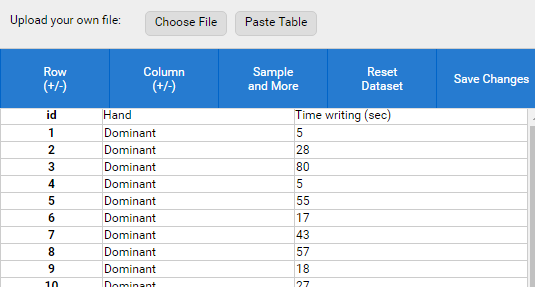
One column will be your qualitative variable (groups), and the second column will be your quantitative variable (measurement or counts).

### Example



### NZGrapher

The next step is to open NZGrapher and **Paste Table** into the programme,

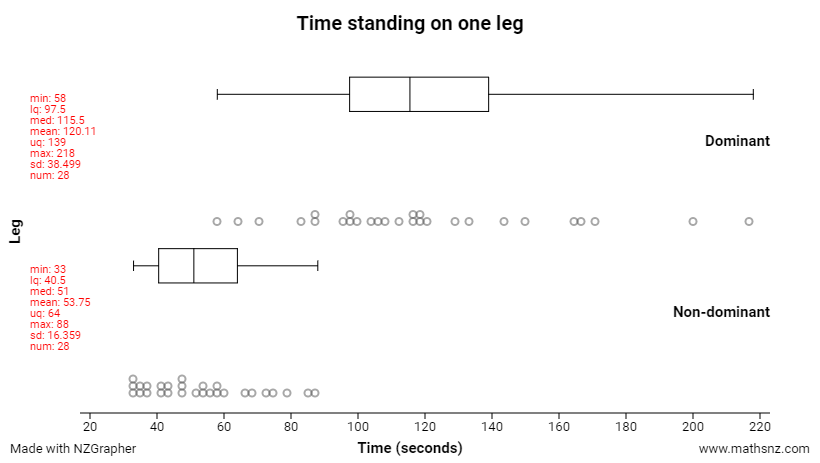


Then you can get the following information (just as you did for the Inference topic):

* Dot plot
* Box and whisker plot
* Summary statistics

Don’t forget to put a title on your graph and label your axis.

### Example



## Observations

During the experiment, the observer (and others) may notice behaviour or events that could affect the results of the experiment. It is important that these are noticed, recorded, and discussed. Improvements for future experiments may also be noted (for Excellence).

### Example

**Observations, effects and improvements:**

I noticed that some students were using their arms to help balance themselves. Other students kept their hands beside their bodies. I think that this could make a difference in the length of time that students are able to stand on one leg, because having your arms outstretched (like gymnasts do when walking on a beam) will help keep your center of mass stable.

If I was to do this experiment again, I would give instructions to students to keep their hands beside their bodies, so that all students have the same conditions, and therefore the results are likely to be only due to the difference in treatment, not due to other factors affecting the results.

## Experiment 1

**Roles:** 

Presenter:

Materials manager:

Data recorder:

Observation recorder:

**Data:**

In Google Classroom a Spreadsheet for entering data is available. All students in the class can open and edit this table (so you can each enter your own data).

Now it’s time to carry out the experiment with the class, following the instructions you created previously, and record your data on the Google Sheet.

**Observations:**

Make notes on student behaviour that might affect the results – e.g. not following instructions, etc.

Next, you need to copy and paste your data into NZGrapher, and get the graphs and statistics that you need.

# Analysis

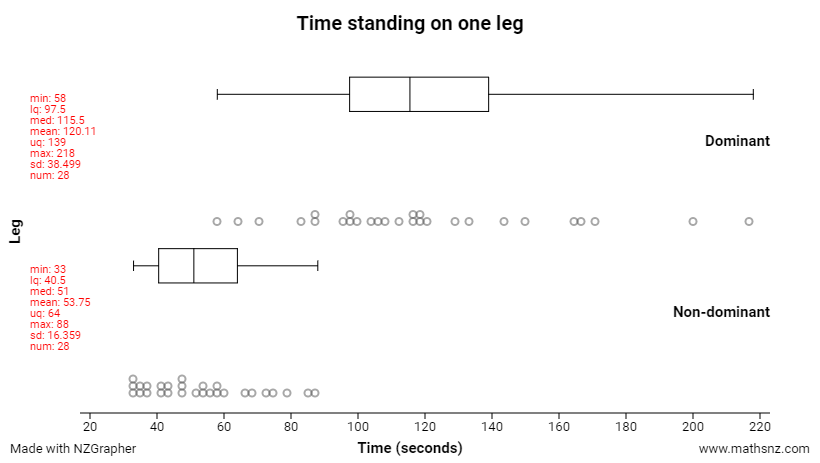
## Features

Here are the features you need to analyse.

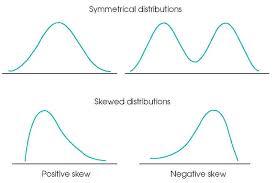
1. **Shape**
2. **Center**
3. **Spread**
4. **Unusual features**

You have covered all these features in the last topic, Inference. If you need any reminders, go back and look in your workbook.

### Example



Describe and justify the features.

**Shape:**

Both the distributions of time that STA3 students can stand on their dominant and non-dominant leg, are skewed to the right, because they have a longer tail on the right hand side.

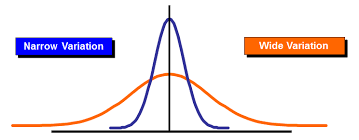
**Center:**

The median time that STA3 students can stand on their dominant leg is around 64.5 seconds longer than the median time that students can stand on their non-dominant leg, which is just over a minute. The evidence for this is that the median time students can stand on their dominant leg is 115.5 seconds, and the median time for writing with the non-dominant hand is 51 seconds.

**Spread:**

IQR (Dominant) = UQ – LQ = 139 – 97.5 = 41.5 seconds

IQR (Non-dominant) = UQ – LQ = 64 – 40.5 = 23.5 seconds

The spread of the middle 50% of data for the time that students can stand on their dominant leg is more spread out than the middle 50% of data for the time that students can stand on their non-dominant leg. The IQR for the Dominant leg is about double the size of the IQR for the Non-dominant leg. The evidence is that the IQR for Dominant leg is 41,5 seconds, and the IQR for Non-Dominant leg is 23.5 seconds.

**Unusual features:**

There are no unusual features in this data set.

## Experiment 1

Describe and justify the features of your data:

## Randomisation Test



The randomisation test helps us to decide if we have enough **evidence** that the treatment may cause an effect on the response variable.

Just like with our Inference topic, the key idea is to find out **if we have enough evidence**.

We also want to use the results of the randomisation test to make an inference.

### Example

We will do a re-randomisation test together as a class.

The dataset is from an experiment where a group of young babies were randomly allocated to either a control group, or a treatment group. The age that the babies first walked was recorded. The control group carried on as usual and did nothing differently. The treatment group had a set of specific exercises to do each day with their baby, and the hope was that the exercises might encourage babies to walk at an earlier age.

Here is the data:

|  |  |
| --- | --- |
| **Group** | **Age walking (months)** |
| Exercise | 9 |
| Exercise | 9.5 |
| Exercise | 9.75 |
| Exercise | 10 |
| Exercise | 12.5 |
| Control | 10.25 |
| Control | 11.5 |
| Control | 12 |
| Control | 13.5 |
| Control | 11.5 |

Is it possible that these babies’ walking ages have nothing to do with whether they undertook the exercises or not? In other words did it matter what group the babies were in or would they have had the same walking age anyway? Is it possible that what we are seeing is just luck of the draw? There are two possible explanations.

1. The data provides **evidence** to suggest there is a **link** between the exercises and walking age.
2. The difference between the walking age could have been produced by **chance alone**.

The key phrase here is **‘chance alone’**, in other words they just ended up this way by random chance.

To find out if it is actually making a difference we want to do a re-randomisation test (this is a bootstrap method, similar to the one we used in the Inference topic). This is done by getting the data and randomly assigning it to one of the two groups.

### Exercise

Instead of us doing the bootstrapping method manually (by cutting out the pieces and randomly choosing them), we are going to go to the website below, and simulate this online. As we do so, you will add the Difference in the medians onto the axis below. We will do this 30 times, rather than the 1000 that the NZGrapher will do for us.

Here is a link to the website we will use: <http://www.jake4maths.com/excon.php>

**Time to First Walk**



Difference Between Medians (Control – Exercise)

The graph that we see above is the Re-randomisation of the difference in medians between the control and exercise groups.

The randomisation test is then added to this graph. Some of you may remember probabilities from a Normal distribution for the Year 12 Probability External exam last year.

## Randomisation test evidence

We want to know if the 2 medians are far enough apart from each other that we can be reasonably sure that the difference is real, and not just random chance.

You need to decide if you have enough evidence to say whether or not there is evidence that the difference in medians is possible **by chance alone**.

**Evidence:**

We look at the probability that has been calculated on the randomisation test.

**The cutoff point is 0.1 (or 10%).**

* If we have a value smaller than this, then we **do** have **enough evidence** that the difference between the medians that we see are **not due to chance alone**.
* If we have a value of 0.1 or higher, then we **do not** have **enough**, and that the difference between the medians that we see **may be due to chance alone**.

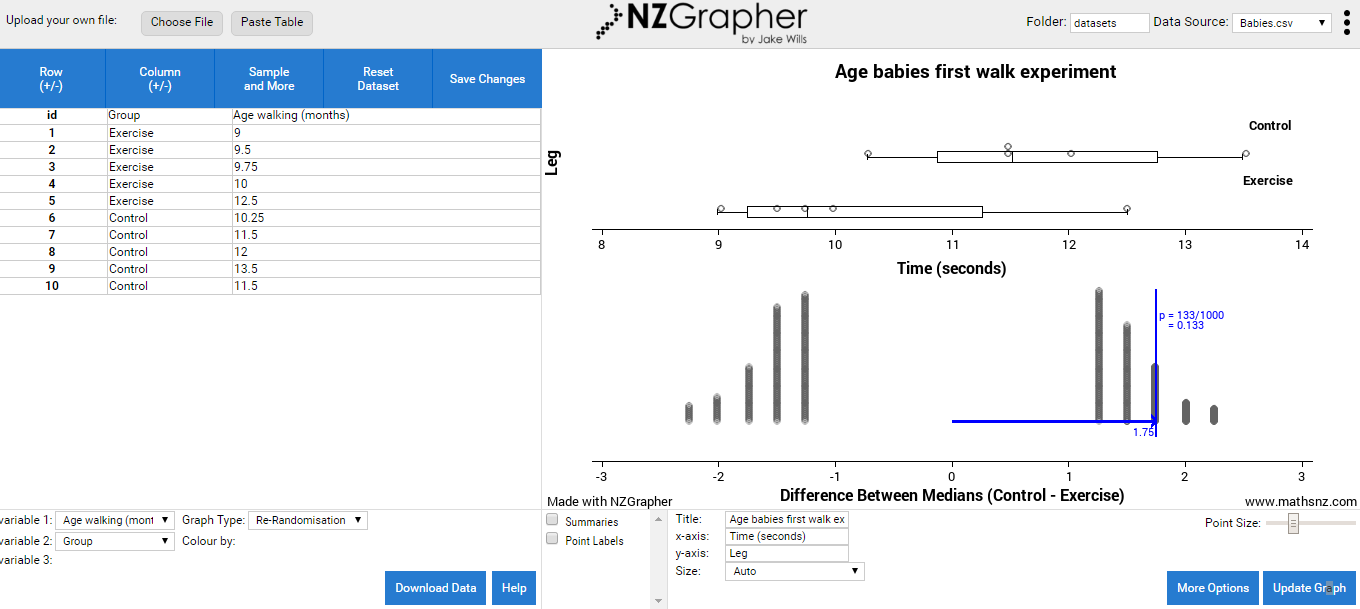




## NZGrapher

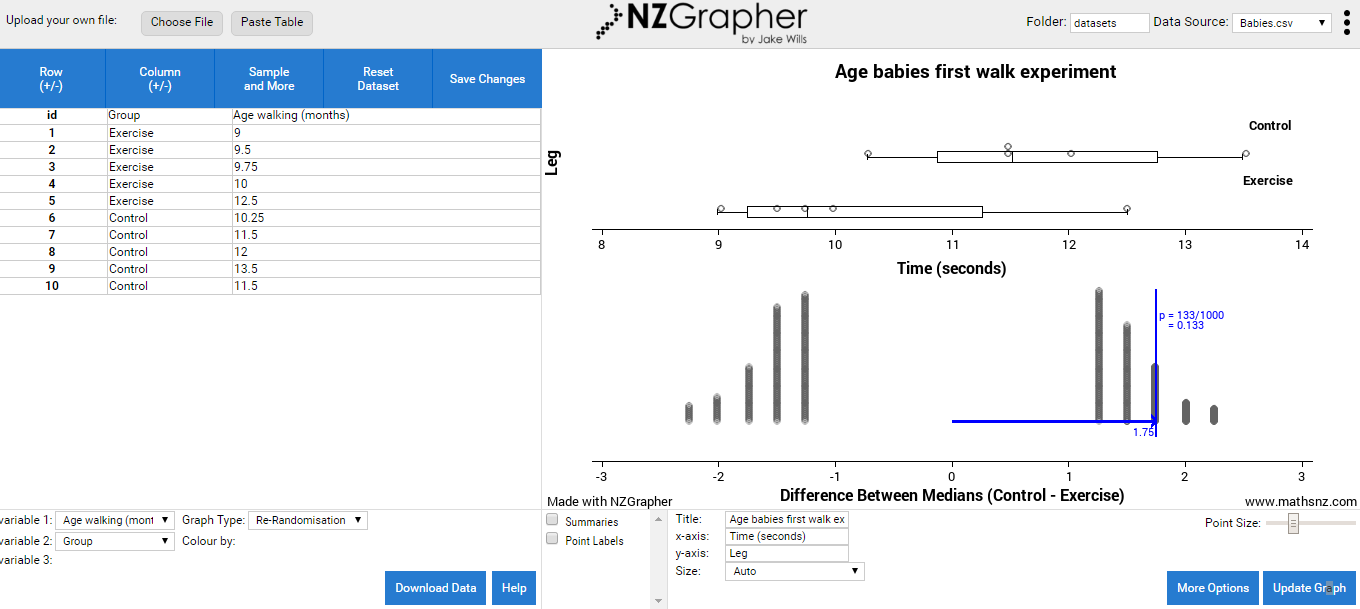
Here is how we create the randomisation test in NZGrapher.

Change the graph type to “Re-randomisation – median”

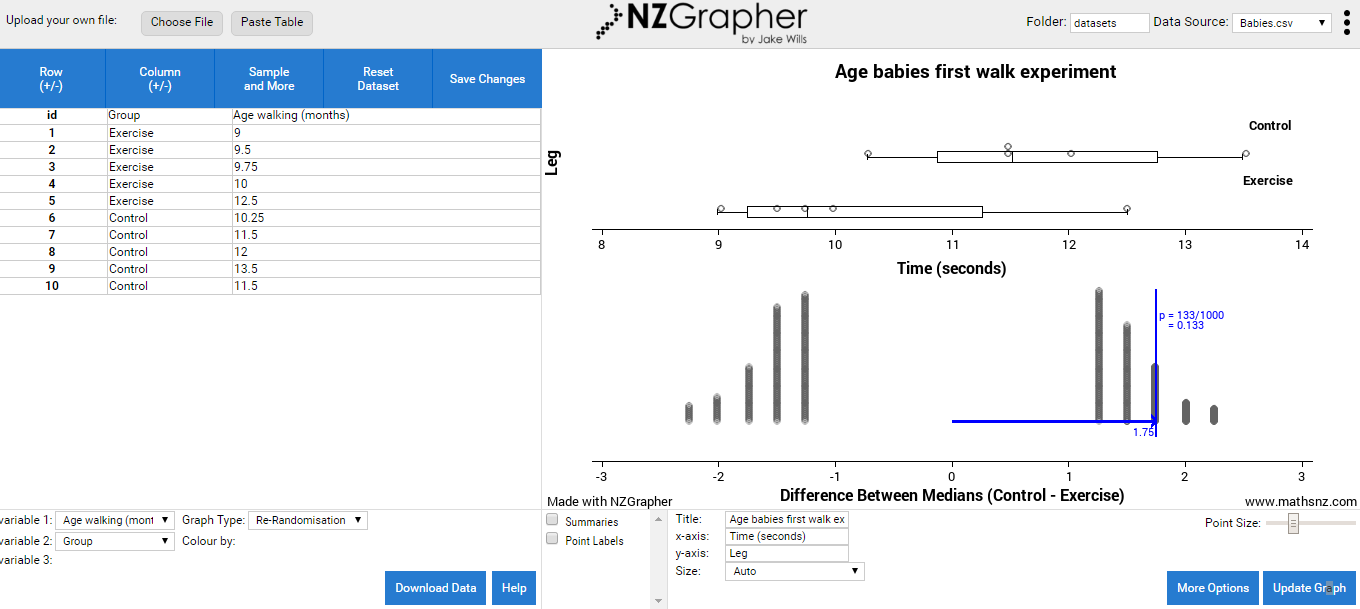


The top part of the graph is identical to the dot plot and box and whisker plot you created previously.

The bottom part of the graph is the bootstrap re-randomisation of the median difference in walking age between the control and exercise groups.

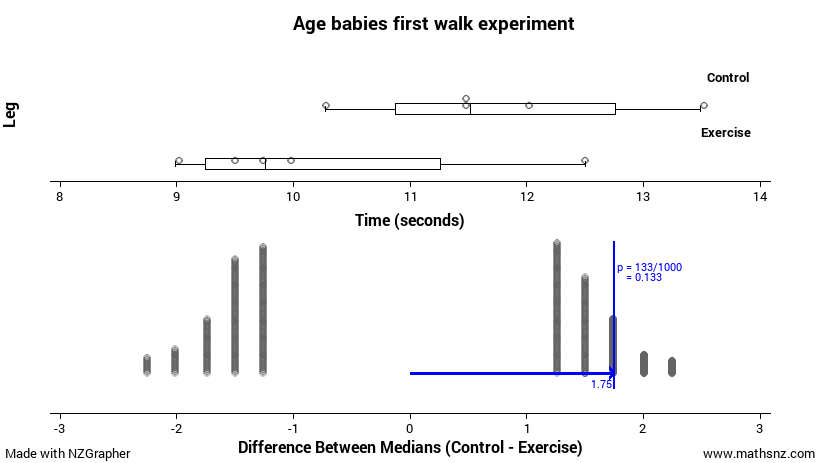


The randomisation test refers to the probability found on the right hand side of the bootstrap distribution graph.



### Example

Here is the bootstrap distribution and randomisation test for our babies walking data.

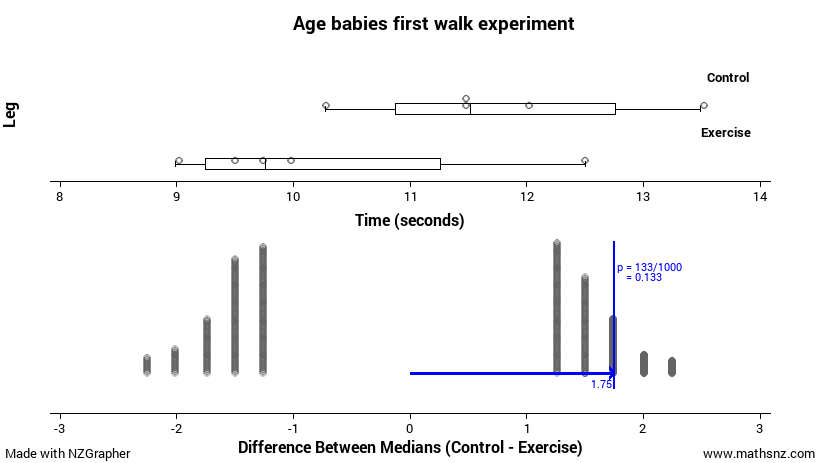


The bottom part is the randomisation test, and we first want to look at the blue arrow, with 1.75 written underneath.

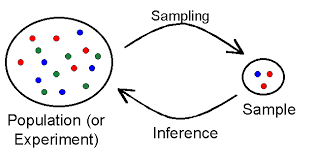
This arrow represents the difference between the two medians, Control – Exercise. Because the difference in medians is positive, this tells us that the Control group median walking age is bigger than the Exercise group median walking age, by 1.75 months.

I want to find out if the difference between the medians of 1.75 months is likely to be just from chance alone, or if this could be due to the exercise program having an effect on when babies first walk.

I used the randomisation test on the medians, and the results are shown below:



**Inference:**

A difference of 1.75 or more comes up 133 times out of the 1000. This is a probability of 0.133 or 13.3%.

As 13.3% of estimates produced by random allocation are **at least as far from zero** as the observed estimate, then the data provides **no evidence** that there may be a link between the two variables. This means that because the probability is above the 10% cutoff point, it would be **likely** that a difference of 1.75 months could happen **by chance alone**.

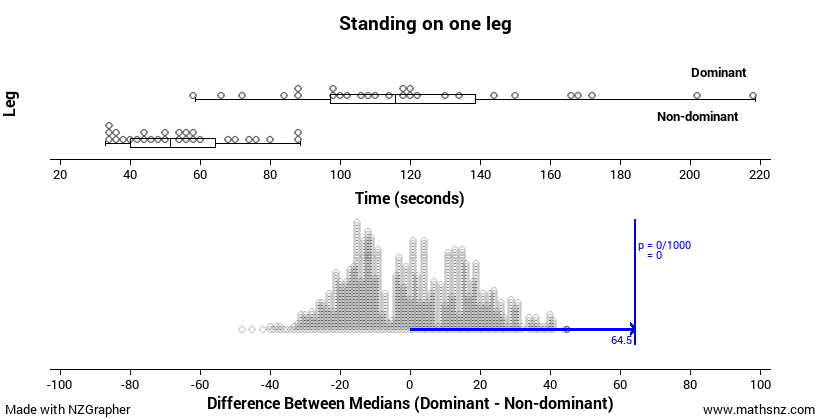
These results apply only to the babies tested in this experiment, but as they were randomly allocated to either the treatment or control group, it is possible that these results may apply to other babies, similar in characteristics to these babies.

## Example

The arrow represents the difference between the two medians, Dominant – Non-dominant leg. Because the difference in medians is positive, this tells us that the median length of time students could stand on their dominant leg is bigger than the median length of time students could stand on their non-dominant leg, by 64.5 seconds.

I want to find out if the difference between the medians of 64.5 seconds is likely to be just from chance alone, or if this could be due to the effect of standing on your dominant leg.

I used the randomisation test on the medians, and the results are shown below:



**Inference:**

A difference of 64.5 seconds or more comes up 0 times out of the 1000. This is a probability of 0.0 or 0%.

As 0% of estimates produced by random allocation are at least as far from zero as the observed estimate, then the data provides **evidence** that there may be a **link** between the two variables. This means that because the probability is low, it would be **highly unlikely** that a difference of 64.5 seconds would happen **by chance alone**.

These results only apply to the students at Ormiston Senior College, from the STA3 class. It is possible that these results may be extended to apply to all Year 13 students at Ormiston Senior College, as it is likely that the students in this experiment are representative of all Year 13 students at Ormiston, and therefore similar in characteristics such as age, height and weight.

## Experiment 1

Make an inference about your data.

# Conclusion

In your conclusion you need the following:

* To answer your investigation question
* Justify your decision based on evidence from the randomisation test.

## Answering the investigation question

Our investigation question is in the form:

I wonder if this treatment **effects** the response variable, for our experimental units.

In order to answer this question, we need to know if we have enough evidence to support this. 

**Evidence:**

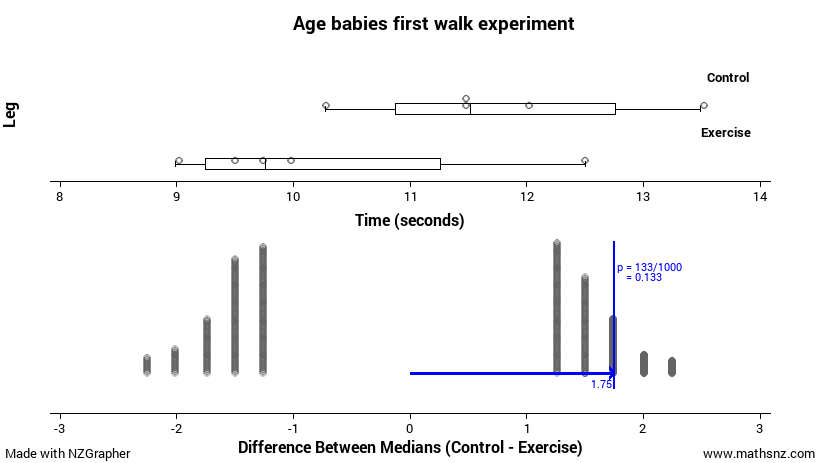
We look at the probability that has been calculated on the randomisation test.





### Example

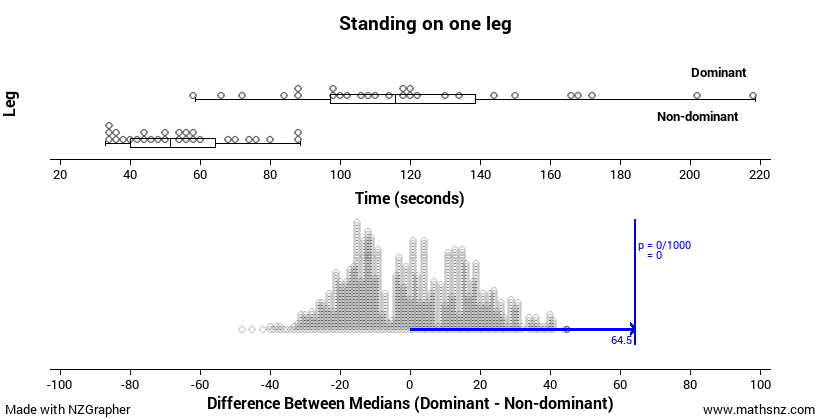
Looking at the randomisation test for our babies data:





Overall I see that there is **not enough evidence** to support the claim that babies who exercise are likely to walk before babies who do not exercise.

### Example



Draw a conclusion for your experiment: 

Because the tail proportion is **less than 10%,** this means that the observed difference in the median length of time year 13’s balance on their dominant or non-dominant leg is **unlikely when chance is acting alone**.

Because the students were **randomly assigned** to the treatment and control group (standing on the non-dominant or dominant leg respectively), we have **enough evidence** to claim that standing on your dominant leg **is effective** in increasing the time that students can balance on one leg.

## Experiment 1

Draw a conclusion about your data.

## Sample Internal (at Achieved level)

|  |  |
| --- | --- |
| **Attractive People** | Title is given |
| **Problem** | |
| It is often stated that attractive people look younger and that smiling makes you look more attractive, and being younger and more attractive is something that we all want. | Reason given for investigation |
| Therefore I wonder if whether or not a person is smiling or not causes a difference in the estimate of the person’s age based on showing a photo of the same person smiling and not smiling to a random sample of students in a year nine class. | Causal relationship question posed |
| I predict a person who is smiling will look younger than someone who is not smiling. | Prediction given |
| **Plan** | |
| For the experiment I randomly selected a Year 9 Science class and flipped a coin to determine if they were showed a photo of Robert Downey Jr. smiling if they got a heads or him with a straight face if they got tails. I then asked them how old they thought he was and they responded verbally which I recorded on my sheet in two columns, smiling and not smiling. | Experiment is described including how groups chosen |
| The treatment variable is if the person in the photo is smiling or not, and the response variable is the person’s age in years. | Identification of treatment and response variables |
| In order to reduce any variation due to other factors, both of the photos were taken of him in the same outfit (his outfit at the People’s Choice Awards in 2013) so they are also both of him at the same age (48). | Other sources of variation |
| **Experiment** |  |
| A possible issue is that because the test was done verbally I may have miss heard the responses or the students may have heard what others were saying. | Any issues that arose stated |
| **Data** | |
| http://www.jake4maths.com/grapher/imagetemp/Dotplot-kKEzYbxemg.png | Graph displayed with summary statistics |
| **Analysis** | |
| The age estimates of the photo of Robert Downey Jr. smiling were much more spread out than the photo of him with a straight face as shown by the width of the box and whisker plot. The smiling group’s median was 0.5 years smaller than the not smiling group, and the means only had a difference of 0.416 years. The data from the groups look like what I would expect to see if chance was acting alone in the experiment. | Key features of the displays and statistics are described |
| I need to find out if a difference between the medians of 0.5 years is likely to just be from chance alone, or if this could be due to smile in the photo. I used the randomisation test on the medians and the results are shown below. | Statement of what test is going to be carried out |
| http://www.jake4maths.com/grapher/imagetemp/BSRerandomMed-GLknV14eJy.png | Results of the test are displayed |
| A difference of 0.5 or more comes up 457 times of the 1000. | Summary given |
| As 45.7% of estimates produced by random allocation are at least as far from zero as the observed estimate, then the data provides no evidence of a link between the two variables. This means that because the probability is high, it would be highly possible a difference of 0.5 years could happen by chance alone. | Inference stated |
| **Conclusion** |  |
| Overall I see that there is not sufficient evidence to support the claim that people who are asked to estimate the age of a person who is smiling will give a different estimate than those who are asked to estimate the age of the same person who is not smiling. | Conclusion given |
| **Appendix: Data** |  |
| Estimates of ages:  Straight Face: 36, 43, 35, 32, 38, 53, 45, 42, 40, 39, 38, 50  Smiling: 53, 39, 38, 36, 37, 43, 30, 39, 50, 35, 45, 51 | Raw data given as appendix. |

**Level 3 Experiments marking grid**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Achieved** | | **Merit** | | **Excellence** | |
| **Problem** | Causal investigation question stated. |  | Research supports the investigation question. |  | Research integrated throughout the purpose drives the investigation question. **Compulsory.** |  |
| Purpose stated. |  | Prediction given. |  |
| **Plan** | Experimental units stated. |  | Justify the allocation of the treatment to the experimental units. |  | **Reflection** on the plan.  This may include how to control sources of variation, blinding (single or double), or refining the definition of the treatment and response variables. |  |
| Describe treatment variable |  |
| Describe response variable |  |
| Describe random allocation of the treatment to the experimental units. |  |
| Describe data collection and recording methods |  | Justify how a source of variation may affect results. |  |
| Describe one possible source of variation (controlled condition). |  |
|
| **Data** | Data collected. |  | Discuss how an observation might affect the results. |  | **Reflection** on improvements to the experiment. |  |
| Observations recorded. |  |
| **Analysis** | Displays:   * Dot plots * Box and whisker * Randomisation test * Summary Statistics |  | Inference is justified, using the strength of evidence for the causal relationship. |  | **Reflection** on the inference or randomisation distribution, showing statistical insight.  This may include a discussion on accuracy of the inference, or insight into the randomisation distribution. |  |
|
| One feature discussed in context (shape, center, spread). |  |
| The randomisation test is used to make an inference about the causal relationship. |  |
| **Conclusion** | Answers the investigative question, correctly interpreting if there is sufficient evidence. |  | Conclusion justified and linked to the research and purpose. |  | Conclusion integrates research and findings.  This may include a comparison of the purpose with findings from the experiment, or a generalisation of results. **Compulsory.** |  |
|
|
|  | **All required** |  | **All required** |  | **Compulsory and one reflection** |  |