National Certificate of Educational Achievement
TAUMATA MĀTAURANGA Ā-MOTU KUA TAEA

Exemplar for Internal Achievement Standard

Mathematics and Statistics Level 3

This exemplar supports assessment against:

Achievement Standard 91583

Conduct an experiment to investigate a situation using experimental design principles

An annotated exemplar is an extract of student evidence, with a commentary, to explain key aspects of the standard. It assists teachers to make assessment judgements at the grade boundaries.

New Zealand Qualifications Authority

To support internal assessment

	Grade Boundary: Low Excellence
1.	For Excellence, the student needs to conduct an experiment to investigate a situation using experimental design principles, with statistical insight. This involves integrating statistical and contextual knowledge throughout the investigation process, and may include reflecting about the process, discussing how possible sources of variation were dealt with during the design phase; considering other relevant variables. This evidence is from a student's response to the TKI task 'Tricky questions'. The student has integrated contextual knowledge gained from research when posing an investigative question about a given experimental situation (1) and when designing the experiment (2). The student has also discussed how possible sources of variation were dealt with during the design phase (3) and made an appropriate formal statistical inference by assessing and interpreting the strength of evidence about the causal relationship (4). When communicating their findings, the student has considered other relevant variables that may contribute to variation (5) and in their conclusion, they have reflected on the process by discussing the possible implications of increasing the value of the anchor (6). This extract is from a student response which also included evidence of selecting and using appropriate displays and summary statistics to an appropriate level for the award of Excellence. For a more secure Excellence, the student could have reflected on the process more fully by discussing the effect of the value of the anchor relative to the true population of Venezuela. The student could also have explained how the results of the experiment were consistent with the research studies discussed when posing their question.

Student 1: Low Excellence

NZQA Intended for teacher use only

I investigated possible bias in questionnaires from using anchors. I wanted to find out if you could get people to give higher answers for a question by using an anchoring question before it. This would be important to take into account when using questionnaires to collect data so you don't unknowingly influence answers, or something that might be used in questionnaires to trick people into giving a certain answer.

I spent some time researching each of the different types of questionnaire bias given in resource A, and it seemed to me that anchoring effect would be interesting to investigate because of the applications of this to everyday life. In my research, I found examples of how 1
people use anchoring bias when selling cars to persuade people into paying more. The research about anchoring bias shows that when people are uncertain about something, they use whatever information is available to help them decide, even if the information is not valid or reliable. I wanted to do a similar experiment, so my investigative question was "will higher estimates be given for the number of people who live in Venezuela if a larger number is used for the anchor?" I used Venezuela for my experiment as I thought that it was a country not many people would know exactly how many people lived there, and for the experiment to work people have to be unsure about the answer or amount for the anchor to have an effect. The research suggested that the higher the number I used in the anchor, the higher the estimates of the number of people who live in Venezuela would be, and so this is what I expected to find in my experiment.

I used a comparison of two independent groups for my experiment, where one group was given a high anchor and one group was given a value close to the real value. I used 57 Year 13 students. I used single blinding, where the participants didn't know which treatment they were getting. In fact, I concealed the fact that it was an experiment at all by presenting the questionnaire as a general knowledge survey.

The response variable was the estimate for the number of people in Venezuela (in millions).
The treatment variable was the number used for the anchor. I had two treatment groups: for 2
one group the anchor was 60 million, for the other group the anchor was 30 million. I used these two numbers as the population of Venezuela is around 29 million.

I created two different questionnaires for my experiment. I included an introduction for the questionnaire used in the experiment that said it was a general knowledge survey, and asked a couple of other questions in the questionnaire so that people would not guess the point of my experiment. In my questionnaire, I decided to ask people to estimate of the number of people in Venezuela to the nearest million because I was confident this would still give me enough variability in the estimates and it would be easier for people to answer the question. The two versions of the questionnaire are exactly the same, except for the number used for the anchor in the question before the one that asks people to estimate the number of people in Venezuela. For this experiment, it was important that people didn't realise there were two different versions of the questionnaire. Before we gave them to students to fill out, we turned the questionnaires upside down and thoroughly shuffled them into a pile to hand out. This was done to make sure that each student did not know which of the two questionnaires they were getting. In this way, we would be randomly allocating students to one of the two treatment groups when the teacher gave out the questionnaires to complete.
There are some factors (identified from my research) I couldn't control for my experiment:

- whether people already know the population of Venezuela (maybe people who had been travelling or international students from South America)
- whether people would take the survey seriously and not give silly answers 5

- whether people were aware of the anchoring bias (which maybe students doing psychology might know about)
- whether some people had better general knowledge than others

By randomly assigning people to one of the two treatment groups, I attempted to balance the possible effect of these variables on the estimates across the two groups.
The variables I controlled for my experiment were:

- giving the same instructions to people about completing the questionnaire
- both groups doing the experiment at the same time of the day (
- same test conditions used for completing the questionnaire
- an independent person carrying out the experiment (the teacher of the class) so that I didn't influence the results if I was the handing out the questionnaires 3

Rerandomisation

I found the median estimate of the population of Venezuela for each group and the difference was 19 million. Using the result from the rerandomisation test, I have very strong evidence that the use of an anchor of 60 million would cause estimates that tend to be higher than when a 30 million anchor is used. This is because when I compared the observed difference between the group medians (19 million) to the distribution of re-randomised differences, a difference of 19 million or higher came up only twice in a 1000 re-randomisations. This shows that it would be very unlikely that a difference as large as 19 million could happen just by chance. It is this test result that provides me with the very strong evidence that 4
chance was probably not acting alone in this experiment but something else, namely the anchor effect, was acting along with chance to create the observed difference of 19 million.

I thought that the experiment turned out reasonably as planned, although not everyone completed the survey correctly, so I couldn't use all the results (see my notes in the appendix). However, there weren't many incomplete or invalid responses, so this shouldn't have affected my data too much. I could have had the teacher check the questionnaires as they were handed back so that responses could be clarified.
The result from the randomisation test, the fact that my experiment was well designed and executed means that I can claim that, for this group of students, an anchor of 60 million is likely to cause estimates that tend to be higher than when a 30 million anchor is used. My results, if we can widen them to beyond this group of students, are important in terms of how they apply to questionnaire design, and the importance of making sure that anchors are not used that may influence people's answers to questions in the questionnaire.
I wonder how much the numbers I used for each of the treatment groups (the numbers for 6
the anchors) affected the estimates? When I looked at the data for the two treatment groups, it seemed that people were not confident estimating the number of people in Venezuela to be as high as 60 million (the median estimate was 43 million for the group with this anchor). Maybe if I had used a value like 100 million they would have ignored it because it would have been unrealistic. Would I have got the same result (my conclusion that a larger number for the anchor would result in (cause) estimates that tend to be higher) if I used a high anchor of 40 million? So perhaps anchors can influence people's responses to answers, but only if they are a certain value.

	Grade Boundary: High Merit
2.	For Merit, the student needs to conduct an experiment to investigate a situation using experimental design principles, with justification. This involves linking components of the process of investigating a situation by experiment to the context, explaining relevant considerations in the investigation process, and supporting findings with statements which refer to evidence gained from the experiment. This evidence is from a student’s response to the TKI task ‘Estimation’. The student has linked the development of a suitable investigative question about a given experimental situation (1) and the design of the experiment (2) to the context. They have explained relevant considerations in the design phase (3) and when managing possible sources of variation (4). The student has also made an appropriate formal statistical inference which has been justified (5) and in their conclusion, there is evidence of linking findings to the context (6). This extract is from a student response which also included evidence of selecting and discussing appropriate displays and summary statistics to an appropriate level for the award of Merit. To reach Excellence, the student would need to integrate the information found in their research into the context in both their introduction and conclusion. The student could also consider in more depth other relevant variables that may have contributed to the variation.

.

Student 2: High Merit

NZQA Intended for teacher use only

My experiment is about guessing the age of a person by looking at the picture of a person. I will investigate whether wearing make-up has an effect on how people 1
guess the age. In my research I have found that celebrities wear makeup aimed at making them look very much younger than their real age and making it harder to guess their age. From the information/ideas found from research for this experiment, it shows that comparing two photos, one with makeup on and another one without make up, can vary the estimated age of a person. The photos of a person with makeup and another without make up can lead people toward an answer which is different, because makeup affects the look of a person by covering dark and red 2
spots, wrinkle, discoloration area, breakouts and any other undesirable spots or areas on their face. These things make photos of a person with and without makeup guide people into different estimates of the real age. I'm not sure if wearing makeup will always make people look younger, because often people wear makeup to look older (like teenagers), so my problem for this investigation is "Does changing the picture of a person wearing makeup and without makeup have an effect on the guesses of the celebrity's real age?"

I will investigate this problem by using Kim Kardashian photos that will be found from Internet for students to estimate; by the way her real age is 38. There are two treatment groups in this experiment: one is the photo of Kim with her makeup on and another one is the photo of Kim without her makeup, to see if there is a difference of the estimation of her age between the two photos/groups. These photos have a question "How old is the person in the photo? " underneath.

How old is the person in the photo?

How old is the person in the photo?

They will be printed and cut individually. There are 50 photos of Kim altogether, 25 identical photos with makeup and 25 identical photos without makeup. The experimental group will be year 13 students a total 50 of because there where two maths classes on at the same time that had a total of 50 people. On the combined class list I randomly assigned each student a 1 or a 2 using my calculator. If the student was a one they were going to be given a picture of Kim without make 3
up. I had to stop giving students ones when a got to a total of 25. A two meant they were going to getgiven a photo of Kim with makeup.

At the start of class the teachers helped me tell the students which room to go to.

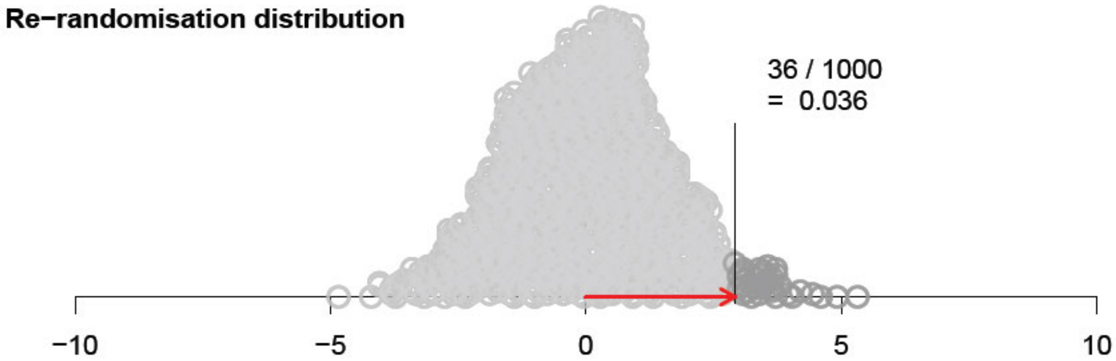
The students were sent to the two classrooms. All the students with a 1 went into one class and all the students with a 2 went into another class. I then went into the first class and carried out the experiment giving out the picture without makeup. None of the students were told it was Kim Kardashian in the photo. When they had finished I asked them to raise their hand and I collected the photos with the answer to the question back. I then went into the second class and repeated the process with one exception that the photo this group saw was one of Kim Kardashian with make-up. I will record the data from the collected answers on to a spreadsheet. The response variable of this experiment will be the estimated age of Kim Kardashian in years.

The variables I could control in this experiment were

- Each class answered the question in the same test condition, on the same date.
- The students who were ones and twos got the same photos of Kim Kardashian, which was separated into two treatment groups of with makeup and without makeup. 4
- I didn't tell either group who was in the picture.
- The photos that the students got, had the same question "How old is person in the photo?" underneath.
- I was the one handing out the photos to the students and collecting them and I was the person who recorded the data.

The variables that I could not control are:
- The personal knowledge of the students because some students could have recognised the person in the picture and might know the age of Kim Kardashian from magazine, Internet or TV shows.
- Some of them might have just guessed random answers because they weren't being serious.

From my analysis I discovered that the difference in the mean ages estimated by each group was 2.92 years and the group given the photos with makeup on guessed a lower mean age.
I then looked for evidence to answer my investigative question. I need to find out if it is likely to get a difference as big or bigger than 2.92 by chance alone. I used the randomisation test 1000 times to produce just by chance 1000 differences between the group means. The graph and results produced from this method are shown below:



My observed difference of 2.92 years only came up 36 times out of 1000.

As the estimates produced by random allocation of 3.6% are at least as far from zero as the observed estimate, then the data provide some evidence of a link between the two variables. This means that because the probability is low, it would be unlikely that a difference of 2.92 years could happen by chance alone, so something else must be working with chance to explain the effect. I can therefore make a call that Kim Kardashian wearing make-up did cause the students to guess Kardashians age.to be lower than it actually is. 5

My results are also important in terms of how old make-up can make you look.
If I had a chance to do this experiment again I would use photos of an ordinary person when wearing make-up and without wearing make-up for students to estimate, to control the variable of personal knowledge because using Kim Kardashian who is a celebrity, most of people might already have some knowledge about her, making this knowledge a variable that we cannot control. So therefore using someone so well known such as Kim Kardashian, where teenagers are more than likely to know her actual age may have an affect on the outcome of the experiment. 6

	Grade Boundary: Low Merit
3.	For Merit, the student needs to conduct an experiment to investigate a situation using experimental design principles, with justification. This involves linking components of the process of investigating a situation by experiment to the context, explaining relevant considerations in the investigation process, and supporting findings with statements which refer to evidence gained from the experiment. This evidence is from a student's response to the TKI task 'Tricky questions'. The student has linked the development of a suitable investigative question about a given experimental situation (1) and the design of the experiment (2) to the context. They have explained relevant considerations in the design phase (3) and when managing possible sources of variation (4). In their conclusion the student has made an appropriate formal statistical inference with their findings being supported with evidence gained from the experiment (5). This extract is from a student response which also included evidence of selecting and using appropriate displays and summary statistics to an appropriate level for the award of Merit. For a more secure Merit, the student could strengthen the justification of decisions made in planning the experiment. For example, by providing a more detailed description of how they managed sources of variation and identifying the variables that cannot be controlled. The student could also provide stronger evidence of linking the research to the situation being investigated and to the findings.

Student 3: Low Merit
NZQA Intended for teacher use only

I found this article online about anchoring effects:
http://www.overcomingbias.com/2007/09/anchoring-and-a.html

Paragraphs from the article:

- Suppose I spin a Wheel of Fortune device as you watch, and it comes up pointing to 65. Then I ask: Do you think the percentage of African countries in the UN is above or below this number? What do you think is the percentage of African countries in the UN?
- Tversky and Kahneman (1974) recorded the estimates of subjects who saw the Wheel of Fortune showing various numbers. The median estimate of subjects who saw the wheel show 65 was 45%; the median estimate of subjects who saw 10 was 25%.
- The current theory for this and similar experiments is that subjects take the initial, uninformative number as their starting point or anchor; and then they adjust upward or downward from their starting estimate until they reached an answer that "sounded plausible"; and then they stopped adjusting. This typically results in under-adjustment from the anchor - more distant numbers could also be "plausible", but one stops at the first satisfying-sounding answer. 1

I decided to investigate student's knowledge about the school and whether I can use an anchoring question to influence answers. My question was "Will having a high anchor first influence estimates for the proportion of students who walk to school be higher?"

Because I don't think people will know what the actual proportion of students who walk to school is, I think I should be able to trick them into giving higher estimates when I use a high number for the anchor, as it says in the article. 2

For my experiment I had to choose a response variable that students would not know the exact answer for (they may have an idea about its value but I would still expect variation in the estimates given). I asked students to estimate the proportion of people who walk to school. I made up a short survey about the school, and asked questions like "What year level are you in?" and "How many students are there at the school?" and other questions that looked like the survey was about finding out what they knew about the school.

For the anchors I chose the two numbers 30% and 60%, because around 30% of students walk to school, 60% is double the actual proportion.

Below is part of the survey I used:

You have been randomly assigned a number between 1 and 100. Your number is ______. Do you think the proportion of students at our school who walk to school is above or below this number? Estimate the proportion of students at our school who walk to school. _________

I took this idea from the article, and I hand wrote either 30% or 60% on each survey sheet, to make it look even more like it was a random number (even though I only used 30% or 60%). 3
I made up equal numbers of each version of the survey (with either 30% or 60%).

I went to two different classes (both Year 9 classes). I had a bag with an equal number of red and white balls in it. The students picked a ball out. If it was red they had the high anchor survey, if it was white they had the low anchor survey. This made sure the experimental units (the students) were randomly allocated to the survey.

I made sure that the students do not look at each other's surveys. I also told them it was a personal survey and that I was interested in their response and how much they knew about the school. My experiment was a comparison of two independent groups design, so students only completed one version of the survey. 4
Some students might know the proportion of students who walk to school (if they are involved with travel wise or the school council) but they should be in both of the groups because I randomly mixed up the different surveys before handing out.

The difference between the median estimates from the high anchor and the low anchor was 18.41%.

This could happen by chance just by randomly allocating the people to two different groups, so I need to do the randomisation test to see how many times a difference of 18.41% comes up when the estimates people gave are re-randomised to the two groups (30% and 60% anchor) and the differences of the means of the two groups are calculated.

The difference of 18.41% or higher came up 8 times in 1000.

The design of my experiment was good and I carried it out well, so I am happy that there are no other explanations for what I see in the data (that the 60% anchor group has estimates which tend to be higher than the 30% group, with a difference of 18.41%) apart from chance and the anchor questions I used.

The randomisation test gives me very strong evidence as it shows me that in this experiment it would be very unlikely that a difference as large as 18.41% could happen by chance alone. This means I can claim that the use of the anchoring question had an effect on the estimates 5
for the proportion of students at our school who walk to school, in particular that the higher anchor of 60% caused estimates that tended to be higher than the estimates from the anchor of 30%. I can only claim this for the group of students in the experiment, but it seems reasonable to expect that it would be true for any group of people.

	Grade Boundary: High Achieved
4.	For Achieved, the student needs to conduct an experiment to investigate a situation using experimental design principles. This involves showing evidence of using each component of the investigation process. This evidence is from a student's response to the TKI task 'Tricky questions'. The student has posed an investigative question about a given experimental situation and provided some evidence of researching the context (1). Experimental design principles are shown by the student selecting the experimental units (2), determining treatment and response variables (3), determining the allocation of treatments to experimental units (4), determining data collection and recording methods (5) and discussing how they considered other possible sources of variation (6). In their conclusion the student has made an appropriate formal statistical inference with their findings being supported with evidence gained from the experiment (7). This extract is from a student response which also included evidence of conducting the experiment and selecting and using appropriate displays and summary statistics to an appropriate level for Achieved. To reach Merit, the student could link their prediction of what they think will happen in their experiment to the context. The student should also describe in more detail the experimental plan, particularly the treatment variable and the method of random allocation and link the design of the plan and their findings more closely to the context. When making their inference, the student should also interpret the strength of evidence more clearly, and indicate who the results apply to.

Student 4: High Achieved
NZQA Intended for teacher use only

Will leading questions influence the treatment groups to give a specific answer? So will changing how a question is worded affect the way our data is produced meaning answers will be higher or lower?

According to this website http://www.busreslab.com/index.php/articles-and-stories/research-tips/general-research-tips/leading-questions/ a leading question is one which attempts to guide the person's answer. You are supposed to avoid using leading questions in questionnaires so that you get truthful answers. 1

I investigated whether people will be honest about how many times they used their phone in class. If I remind them that it is against the school rules - I think that they will give lower numbers.
Our experimental units will be two Year 12 physics classes of 58 students. Out of the 58 student 29 students will receive one survey and the other 29 will receive another, at random we will be handing out the survey. 2
One group will receive a survey asking "Even though using a cell phone class is against school rules, how many times did you use your cell phone in class last week?" and another group of students will receive a survey asking "How many times did you use your cell phone in class last week?" 29 students were given the survey which had the leading question about 'school rules' in it and the other 29 were given the question without any mention of 'school rules'. This makes the experiment design one of comparing two independent groups. 4
The response variable for our experiment will be the number of times the student writes down they used their cell phone in class over the last week. We chose the last week for the question so we would get range of answers – if we had just said yesterday, then maybe the answers would only range between 0 and 5 times (one time per lesson). The variables that we can control for the experiment include the following: 3 6

- Same test conditions
- Same time

The students will be given the same time of day to complete the survey. I will be telling both the groups the rules before handing out the survey so students don't copy other student's answers. They will also be told to hand in the survey straight after they have finished answering the questions so they don't change their answers. 5
Variables that we can't control include the following:

- A student's memory is a variable we cannot control because some students won't be able to remember how many times they used their cell phone in class.

The uncontrolled variables will be randomly assigned to the treatment groups to balance them.
The experiment will be conducted in the following way:

- The classes will randomly split evenly in to 2 groups. Two students will be handing out the survey and one student will collect them in.

I found the median number of calls for the group who were not reminded it was against the school rules and the median number for those who were reminded. The difference between the median was 2 and the difference between the means was 1.86.
I used the rerandomisation test with the means. This will re-randomise the answers to the two groups (leading question, no leading question) and record the difference between the means of the two re-randomised groups each of the 1000 times.

A difference of 1.86 or higher came up 332 times out of 1000.

In my investigation I found that the leading question I used wasn't effective in making people's answers higher or lower. I could have got a difference between the means of the two groups this size by chance without me doing anything (just by shuffling up the groups) because the value was 33.2%. 7

	Grade Boundary: Low Achieved
5.	For Achieved, the student needs to conduct an experiment to investigate a situation using experimental design principles. This involves showing evidence of using each component of the investigation process. This evidence is from a student's response to the TKI task 'Tricky questions'. The student has posed an investigative question about a given experimental situation (1), planned the experiment using experimental design principles by selecting experimental units (2), determining treatment and response variables (3), determining the allocation of treatments to experimental units (4), determining data collection and recording methods (5) and considering other sources of variation (6). The student has also made an appropriate formal statistical inference and communicated findings in a conclusion (7). This extract is from a student response which also included evidence of conducting the experiment, selecting and using appropriate displays and summary statistics to an appropriate level for the award of Achieved. For a more secure Achieved, the student, when posing a suitable investigative question about a given experimental situation, needs to provide some evidence of researching the context. In their experimental plan, the student could provide more detailed discussion, such as explaining how the surveys were distributed. The answer to the question should also be described more clearly.

Student 5: Low Achieved

NZQA Intended for teacher use only

Question: Does the order of questions in a questionnaire affect a student's answers to the questions? 1

Plan:

My questionnaire was about how many Facebook friends a person has. I wanted to test whether mentioning a link between Facebook friends and popularity first will affect the number of Facebook friends that participants put down.

To investigate this I will use two Year 13 classes as my experimental units. There will be 53 students, 30 female students and 23 male students who will be split up randomly into two groups. 2

Two surveys have been typed up particularly with question order in mind. One asks if popularity is defined by the amount of Facebook friends a person has then is immediately followed by a question that asks the user to state how many Facebook friends they have. The other reverses this order, asking how many friends a person has first, then asking if the number of Facebook friends defines popularity. 6

The treatment variable for my experiment is whether the popularity question is before or after the question asking how many Facebook friends a person has. Students only complete one survey. 3

The two surveys will be carefully distributed to students in the class, by members of the experiment team in order to control the number of boys and girls who receive each kind of survey removing any possible variables that might arise from the difference in gender. We will do this by randomly allocating the boys to the two surveys and then randomly allocating the girls to the surveys. 4

The response variable from my experiment will be the amount of Facebook friends listed by participants. 3

To ensure that students do not discuss the question order of their surveys or how many friends they think they might have, participants are made to fill out the survey in silence without any communication to those around them. 5

Analysis

In the experiment the difference between the mean number of friends when popularity was mentioned first and second was 97.9 carried out the rerandomisation test with the means. The test will see if the difference I got between the means (97.9) of the two groups is likely by chance.

The difference of 97.9 only came up 55 times in the 1000 re-randomisations. Based on the evidence (the rerandomisation test results) I would say that the order of questions does affect answers. Having the popularity question first caused answers for the number of Facebook friends that tended to be higher than when the popularity question was second. 7

	Grade Boundary: High Not Achieved
6.	For Achieved, the student needs to conduct an experiment to investigate a situation using experimental design principles. This involves showing evidence of using each component of the investigation process. This evidence is from a student's response to the TKI task 'Estimation'. The student has posed an investigative question about a given experimental situation (1), planned the experiment using experimental design principles by selecting experimental units (2), determining treatment and response variables (3), determining the allocation of treatments to experimental units (4), determining data collection and recording methods (5), considering sources of variation (6) and selecting and using appropriate displays and summary statistics (7). To reach Achieved, the student needs to make a correct inference. The student should also strengthen the plan by explaining more clearly how the experiment was conducted. When posing the investigative question about a given experimental situation, the student needs to provide some evidence of researching the context.

Student 6: High Not Achieved

NZQA Intended for teacher use only

We chose to investigate if the size of dots might affect people's estimation of the number of black dots on an A4 piece of paper. 1

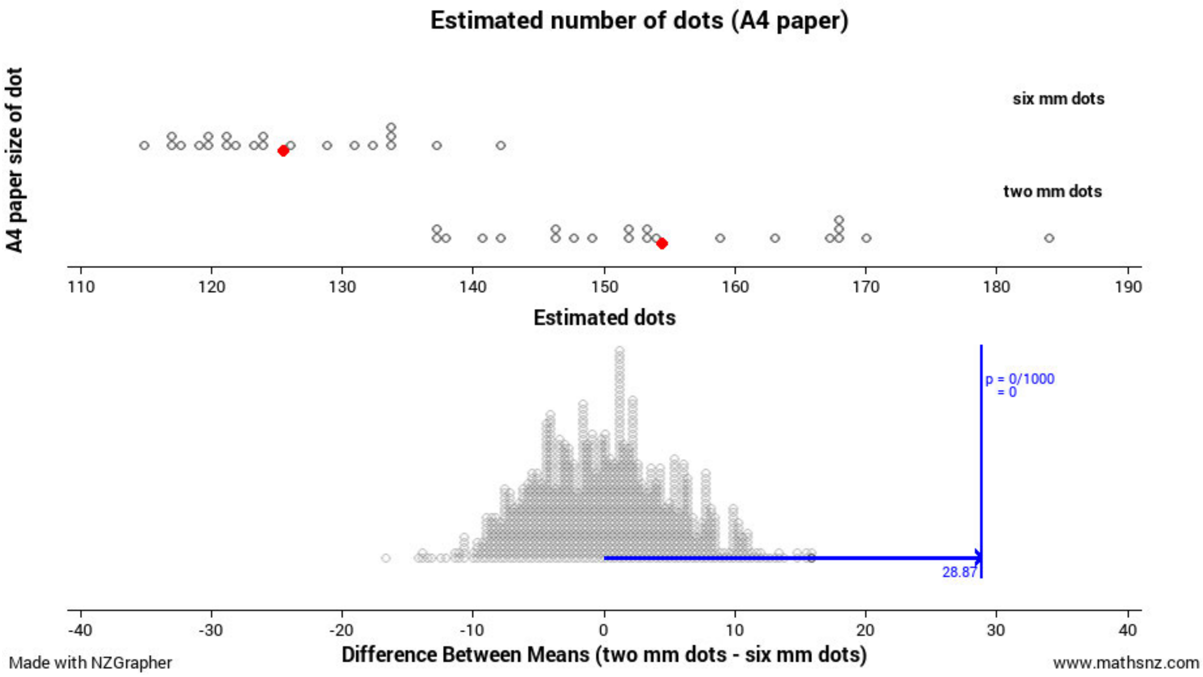
We carried out this experiment by preparing A4 sheets one with very small black dots (diameter of 2mm) and the other with larger black dots (diameter of 6mm). each piece of paper had the dots randomly scattered throughout the page. 3 We did this to try and prevent the people taking part in the experiment, being able to count the dots or come up with some sort of method that they could guess the number the number of dots on the page. 2 The students who did our experiment were two year nine maths classes. We then went into each class the got each student to select a piece of paper one piece had a 1 on it and the other had 2 on it. 4 We then got all the students with 1 on it to go to another room and all the students with 2 on the piece of paper to come into this room. The classes were chosen for us so we could get some silly answers.

The overall design of my experiment will involve a comparison of two independent groups. Students will be given only one of the A4 pages of dots. The student will estimate the number of dots on the A4 page. Each group will be shown the A4 page for 1 minute on a power point slide at the front of the room. and then be given 30 seconds to write down their 5 answer The students will also be told not to communicate to one another while the experiment is running.

A4 paper/ =size of dot	Estimated dots	A4 paper/size of dot	Estimated dots
two mm dots	152	six mm dots	134
two mm dots	163	six mm dots	129
two mm dots	137	six mm dots	137
two mm dots	141	six mm dots	132
two mm dots	153	six mm dots	134
two mm dots	168	six mm dots	122
two mm dots	146	six mm dots	115
two mm dots	170	six mm dots	120
two mm dots	138	six mm dots	131
two mm dots	146	six mm dots	124
two mm dots	152	six mm dots	126
two mm dots	137	six mm dots	121
two mm dots	159	six mm dots	117
two mm dots	168	six mm dots	119
two mm dots	184	six mm dots	121
two mm dots	167	six mm dots	142
two mm dots	148	six mm dots	123
two mm dots	142	six mm dots	118
two mm dots	149	six mm dots	120

two mm dots	168	six mm dots	134
two mm dots	154	six mm dots	124
two mm dots	153		

I did the randomisation test using the means. This will take the values from the groups and randomly re-assign them to one of the two groups and calculate the difference between the re-randomised group medians 1000 times. The results are below:



7

A difference of 2.077 dots came up once out of 1000 for the re-randomised differences.
I can therefore conclude that the size of dots might does not affect people's estimation of the number of black dots on an A4 piece of paper.

My experiment was not designed well. The students could not clearly and confidently estimate the number of black dots on a A4 piece of paper. This was probably due to fact that we had too many black dots on each A4 piece of paper and it was difficult to see the small black dots when it was put onto the power point slide. 6