|  |  |
| --- | --- |
| **Introduction** | Hello. This video is designed for Intermediate teachers to analyse data and draw conclusions with their students using the data viewer on Census as School. We are targeting year 7 and 8 students, working at levels 3 and 4 of the New Zealand Curriculum. It will help if teachers and their students’ have experienced answering the Census at School survey online beforehand. We will use the data detective PPDAC cycle, which is Problem, Plan, Data, Analysis and Conclusion.  So let’s begin. |
| **Problem Plan** | The first stage of the statistics investigative cycle called the PPDAC cycle, is the Problem stage is when you state your statistical investigative question.  Before starting your statistical investigation you will need to understand all the variables used in collecting the data. For example the variable timetravel gives you the time in minutes it usually takes a student to get from home to the school. Reactiontime is the time in seconds it takes for a student to react to a button on the screen changing colour.   |  |  |  |  |  | | --- | --- | --- | --- | --- | | **2013 Questions** | **Variables** | **Options** | **Units** | **Data type** | |  |  |  |  |  | | **Questions about you** |  |  |  |  | | 1. Are you male/female? | gender | male |  | Qualitative | | female |  | | 2. What is your age in years? | age |  | years | Quantitative | | 3. Which country were you born in? | country |  |  | Qualitative | | 4.Which ethnic group do you belong to? | New European | yes/no |  | Qualitative | | Maori | yes/no |  | | Samoan | yes/no |  | | Cook Island Maori | yes/no |  | | Tongan | yes/no |  | | Niuean | yes/no |  | | Chinese | yes/no |  | | Indian | yes/no |  | | Other | yes/no |  | | 5. In how many languages can you hold a conversation about a lot of everyday things? | languages |  |  | Quantitative |   Look at all the variables collected and start thinking of questions you could ask using the data.  For example my investigative question is “I wonder what are typical times spent travelling to school by Year 7 and 8 New Zealand students?”. This is a question that can be answered using the data. Writing a good question is important as all aspects of the investigative cycle depend on this starting point.  The variables of interest are timetravel and year level. These variables are students’ responses to the survey question “How long does it usually take you to get to school? Answer to the nearest minute.” This question was asked in the 2013 and 2011 Censusatschool.  The second stage of the PPDAC cycle is Plan. Since the data have already been collected through the online census my plan is to take a sample of the year 7 and 8 student data. So I need to select the year levels I wish to sample from.  I then need to decide on the size of my sample. We have about 32000 student responses in the census and my sample will be a random selection of let’s say 500 year 7 and 8 students. Right, now I’m ready to get my data. This is the random sample that has been drawn for me to analyse.  Now I have my data I move to the third stage of the PPDAC cycle, where stories in the data can be unlocked from looking at plots of the data or looking at the distribution or shape of the data. |
| **Analysis** | I’m ready to be a data detective and see what stories the data can tell us. Let’s construct a plot of the times students take to travel to school. We need to scroll down and select the timetravel variable. Before I construct a plot of the data, draw a sketch of what the plot might look like. Perhaps collect statistics from the class about their usual travel time to school as a useful beginning. Why might it look like that?   |  |  |  | | --- | --- | --- | |  |  |  | |  |  |  | |  |  |  |   When an appropriate plot/s is constructed by the data viewer, ask students what do they notice? First let’s look at the dot plot. Each dot represents the time in minutes one student takes to get to school. We can see the main body of times is between 5 to 25 minutes. The data are much more spread out at the longer travel times. The shape of the distribution is not symmetrical; it is skewed towards the longer travel times.  I notice that the maximum travel time to school for this sample of students is 120 minutes or 2 hours. Let’s go back and include some summary statistics of the sample and see what the summary statistics tell us. The smallest amount of time year 7 and 8 students take to get to school is 2 minutes according to the summary statistics shown here.  Looking for more stories in the data, you might like to ask the students for ideas about why the mean is higher than the median for this sample.  You might like to introduce the idea of the mean being a point of balance point of the data. Perhaps think of it like the centre of a see-saw. For lesson idea’s see <http://www.nisdtx.org/cms/lib/TX21000351/Centricity/Domain/557/Journal2-2.pdf>    A further discussion you may like to have with your students is, what does the shape of the box plot underneath the dot plot represent? The three features of the box plot you might discuss at Level 3 are the minimum, median and maximum value. At level 4, you might discuss the middle group or middle 50% of the data and the shape or distribution of the data. You can ask students, is the plot symmetrical or not? Is it lopsided or unbalanced and where is the data more spread out, at the longer travel times or shorter travel times? Are there any groups or clusters or extreme values in the data?  Do you believe a student would take 2 hours to travel to school? What might be some reasons for the long travel time? There may not be any extreme values that are very far away from the rest of the data as in this plot.  It is important to discuss the context of the data with the students, because although the data are numbers they are numbers attached to a context.  Who might these students be that spend between 1 and 2 hours getting to school?  Would it take the same time every day?  Let’s look further into the data and explore the story further. |
| **Conclusion** | Now we can answer our statistical investigative question, by saying that typical times seem likely to be in the middle of this main group of data, the data is non-symmetrical which means that many students take 10-20 minutes to get to school, while a few can take up to 60 minutes or more to get to school.  We can also use sample statistics of the mean and median to answer the question about a typical time New Zealand year 7 and 8 students spend travelling to school.  The median tells us the typical time spent travelling to school by these year 7 and 8 students in New Zealand is around 15 minutes. The mean time is slightly higher than the median time at 20 mins. I would think that 15-20 minutes is a typical time spent travelling to school by year 7 and 8 students in New Zealand. This makes sense as most places in New Zealand have a local school that students can attend in their nearest town or suburb. There is a large difference in the shortest and longest times of 2 hours. This means there is a large amount of variation in the times students spent getting to school. This is probably because of the distances students live from school, especially if they live in rural areas and the type of transport that they take to school, school bus or car or biking or walking. Students living in cities might also get held up in traffic.  Another important statistical concept is making an inference about a population from a sample. Ask students if they think that this random sample of 500 students is representative of all year 7 and 8 New Zealand students? Why do they think it is or why might it not be? Can a random sample of 500 students tell us about the population of year 7 and 8 New Zealand student’s? Why or why not?  Continuing the statistical investigative cycle is an important aspect so ask students what else they would like to know about students travel times to school.  I wonder if students who live in the country would spend longer travelling to school on average than students who live in cities? |
|  |  |
| **I wonder…** | Another thing I wonder about is would year 7 students have similar travel times to Year 8 students. I expect they would, as there is only one year difference and I think parents would make sure intermediate students travelled the same way and used the same type of transport for example bus or bike to school each day.  So we now need to separate the data by year level. We should include summary statistics and then have a look at an appropriate plot the data viewer produces. Here the plots are separated into the categorical variables of year 7 and year 8 students. We can see that for this sample of 500 students the median values are very similar/exactly the same? The shape and main group/body of travel time’s look very similar to the last plot between 5-20 minutes. The statistical averages are very similar they tell us that a typical time getting to school for year 7 and 8 students is 15 minutes according to the median and 20 for the mean average time. So this analysis has confirmed my first thoughts about the question. |
| **I wonder…** | I wonder if students who live in different regions around New Zealand have similar travel times to school. I expect there may be differences in the typical times for some regions and you can have a discussion with students what these might be before you begin. So let’s do another analysis but this time separate out the travel times by regions. Now the dot plots show the different regions around New Zealand and the travel times of year 7 and 8 students to school. The box plots show there are differences in the median average travel times of the regions.  The summary statistics show some medians travel times are higher which means longer average travel times to school and some are low, which means students spend as little as \_ mins on average getting to school.  The Northland region has a high median value of \_ mins and mean of \_ mins to school. This is considerably higher than other regions, ask your students why might this happen?  There are a few regions where the data is very spread out, where a few students have over a 50 minute travel time to school. What might be some reasons for this? Would the reasons be the same for an urban region like Wellington or a rural region like Northland?  Who would be interested in this data and your answers to these questions? Have a discussion about what decisions might be made using data about travel times to school. |
| **Contact us.** | I hope you can see how to be a data explorer with your students using the CensusatSchool Data viewer. If you would like anymore help or more movies on data analysis please email us at CensusatSchool. Thanks for watching, see you next time. |