STATISTICS TEACHER
SUPPORTING THE TEACHING AND LEARNING OF STATISTICS
ASA NCTM
STatistics Education Web
STEW
online journal of K-12 statistics lesson plans

A Tale of One City and Two Lead Measurements

Kirk Anderson & Mary Richardson, Grand Valley State University
Published: September 2019

Overview of Lesson

In 2014, a crisis began when the city of Flint Michigan switched its water supply in an effort to save money and the water became contaminated with lead. The EPA's Lead and Copper Rule states that if lead concentrations exceed an action level of 15 parts per billion (ppb) in more than 10% of customer taps sampled, then actions must be undertaken to control corrosion, and the public must be informed. From this, we see that the "action level" of 15 ppb is compared with the 90th percentile of the data. In this lesson, students will (1) read two news stories about the Flint water crisis and answer questions to demonstrate understanding of the articles, (2) compute the 90th percentile using two different datasets (with and without outliers), (3) discuss the computations, (4) use CODAP or other software to create a dotplot, and (5) write a report tying it all together.

Type of Data

- One quantitative variable
- Static dataset provided by lesson plan authors

Learning Objectives

- HSS.ID.A.1 Represent data with plots on the real number line (dot plots, histograms)
- HSS.ID.A.3 Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).

Audience

- Lesson was tested in an undergraduate statistics course for non-majors.
- This lesson may be appropriate for students from grade 6 through an introductory statistics class at the college level.
- *Prerequisites:* Prior to this lesson, students should have experience calculating percentiles (by hand and using technology) and creating graphs to display quantitative data (using technology).

Time Required

45 min of class time + time beforehand for students to read articles + finishing up reports as homework

Technology and Other Materials

- *Technology:* Any tool that can be used to calculate percentiles and create graphical displays.

  This lesson plan uses CODAP – a free, web-based data analysis tool: https://codap.concord.org/ Separate instructions are provided for teachers who prefer to use Excel. It's also possible to modify this lesson so it can be completed entirely by hand.

- Handouts for each group of students

Lesson Plan

Flint is a city in Michigan, with just over 102,000 residents. In 2014, a crisis began when the city switched its water supply to the Flint River in an effort to save money. Due to insufficient treatment, the river water caused serious corrosion to supply pipes, many of which are lead. The tap water in Flint homes appeared brown and smelled bad, and the damage to pipes caused the problem to continue well after the city switched the water source back. This became a national news story, with images of residents picking up cases of bottled water, complaining that they couldn't even use their tap water for bathing. How did the authorities handle the problem? To quote Gov. Rick Snyder, "This was a failure of government at all levels. Local, state and federal officials — we all failed the families of Flint."

The data used in this activity are real and relatable to everyone who expects their tap water to be safe to drink. As students use CODAP or other technology to solve this data investigation, they will be in a position to scrutinize the decisions made by government officials.

Research Prior to the Lesson

Your students will most likely not be experts on the Flint water crisis, so assigning two short news articles as out-of-class readings will give them the needed background to discover how a poor application of statistical concepts played a key role in the story. Additional readings for teachers are listed in the appendix.

- http://michiganradio.org/post/video-how-dropping-two-flints-lead-test-numbers-changed-things-state
- http://michiganradio.org/post/expert-says-michigan-officials-changed-flint-lead-report-avoid-federal-action

Large Group Discussion of the Flint Water Crisis

Motivate the activity with a brief discussion of the Flint water crisis. What follows next is a summary of the key points of the readings, formatted as a Q & A session. You may pose the questions verbally, by writing on the board, or by projecting slides.

Q: Who determines when the level of lead in drinking water is a problem, and what benchmark do they use?
A: The EPA's Lead and Copper Rule gives an "action level" of 15 parts per billion (ppb).

Q: What statistic is used to compare to the action level?
A: The rule states that if lead concentrations exceed 15 ppb "in more than 10% of customer taps sampled," then action must be taken to control corrosion, and the public must be informed. From this, we see that the action level is compared with the 90^{th} percentile of the data, since 10% of values fall above the 90^{th} percentile by definition.

Q: What sample size is required? Is it a random sample?
A: Cities are supposed to test water from at least $n = 100$ homes deemed most at risk. This is definitely not a random sample; rather, the rule seems to recommend a sample biased toward the highest lead levels.

Students know that random samples are the gold standard, but since public health is at stake, they might agree that a convenience sample targeting the homes most likely to have problems makes more sense here.

Q: Did they actually do this (test water from at least $n = 100$ homes deemed most at risk)?
A: No, Flint officials only managed to sample water at $n = 71$ homes, which were not necessarily most at risk of contamination. An expert on water treatment and the ACLU of Michigan obtained documents through a Freedom of Information Act (FOIA) request and found that the city did not test "worst-case-scenario" homes known to either be serviced by lead service lines, or that have lead pipes in them. Even with these 71 lead levels, the 90th percentile was over the action level.

Q: Explain what led to the omission of two lead measurements.
A: After the city submitted their data, the state instructed them to drop two of the values for violating testing guidelines. One measurement was dropped because there was a water filter in the home – which could give an artificially low measurement – but actually this value was 104 ppb, the maximum by far. The other dropped value, 20 ppb, also on the high side, was dropped because it came from a business, not a home.

Q: How did the omission of two lead measurements affect the 90th percentile?
A: Once these values were dropped, the recalculated 90th percentile for the $n = 69$ measurements fell below the action level for lead.

Q: Did the Deputy Director of the MDEQ properly explain how to compute the 90th percentile?
A: No, he said to drop the highest 10% of the data, then take the average. This is a trimmed mean, which is a very different statistic. The correct way to compute the 90th percentile was demonstrated by a professor in the video included in one of the news stories.

Before calculations begin, discuss that there are many ways to compute a percentile. Your students will have already seen one method in the video featured in the first news story. The method that CODAP uses is slightly different from this, and there are others. Why are there so many ways to compute the same thing? A short answer is that before computers, the calculation had to be simple to do by hand, and since then, other methods have been developed that give results which are better in some way. A longer answer can be found at https://analyse-it.com/blog/2013/2/quantiles-percentiles-why-so-many-ways-to-calculate-them.

Hopefully, the video is effective in showing students how to compute percentiles, but you may need to provide further instruction, especially for the interpolation step. We use the first formula (used in the video) on the full ($n = 71$) sample to provide further details on this calculation. The first task is to write out the values in order:

0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 4, 4, 5, 5, 5, 5, 5, 5, 5, 5, 6, 6, 6, 6, 7, 7, 7, 8, 8, 9, 10, 10, 11, 13, 18, 20, 21, 22, 29, 42, 42, 104

You may also want to number each lead measurement from 1 to 71. The next task is to use the index formula with $n = 71$ and $p = 90$: $np + 0.5 = (71)(.90) + 0.5 = 64.4$. This tells us that the 90th percentile is between the 64th and 65th ordered lead level values, which are 18 and 20.

We must "interpolate" between these values. To interpolate means to land on a specific spot in between

the two numbers. If the index had been computed to be exactly 64, the 90th percentile would be 18 ppb. If the index had been computed to be 64.5, we would simply take the average of 18 and 20 to get 19 ppb. In the video, this is not explained very well, and you must pause the video to catch what was written on the board: $18\times(1-.4)+20\times.4=18.8$. A much more intuitive way to interpolate is the following. Since the index number is 64.4, the answer is 40% of the way between 18 and 20. The distance from 18 to 20 is 2, so we compute $2(0.4)=0.8$. Finally, we add that to 18 to get 18.8 for the 90th percentile.

Students Complete Activity Handout

Students may work individually or in groups. Task each student/group with completing the six steps shown on the activity handout. The activity will culminate with a report that summarizes the previous steps of the activity. The full student handout (including correct responses and a sample student report) is attached as a separate document.

(1) Background reading – two news articles

(2) Based on the background reading, students answer a series of questions that are relevant to the data collection procedure used to collect the water samples and why some of the values were omitted by the MDEQ.

(3) Students hand-calculate the 90th percentile for the Flint water data. Two different formulas will be used. The calculations will be performed for the full ($n = 71$) and altered ($n = 69$) samples, for a total of four calculations.

(4) First students are asked to consider whether the choice of percentile formulas was important. Although the formulas give different values for the 90th percentile, they aren't drastically different, especially when we compare them to the 15 ppb action level. Students are often dismayed to learn that there is more than one correct way to compute something, but they will agree that an arbitrary formula choice wasn't the issue here.

Students then use their calculations to explain the effect of omitting the two lead level measurements in question. Students are encouraged to write freely here for at least one paragraph, taking the information from the news articles and connecting that to what they have heard about the Flint water crisis.

(5) After their discussion of how much of a difference two observations can make, students enter the data into CODAP or another software package. The software is then used to compute the 90th percentile, confirming hand calculations. Additionally, dotplots that spotlight the two controversial measurements are produced. The graphs are used by students in a written conclusion that helps illustrate their argument.

Note: If you decide to use a different type of software (or no software at all), you will need to re-write the instructions for this step on the handout. Even if you use CODAP, it's a good idea to double-check that the instructions work for the most up-to-date version.

(6) Students write a report summarizing their work in the previous steps.

Attached Materials

- Student activity handout with correct responses and sample student report
- Instructions for calculating percentiles and creating histograms in Excel

Appendix

Additional Readings about Flint Water Crisis

- To get a basic grasp of the story, read the Wikipedia entry at https://en.wikipedia.org/wiki/Flint_water_crisis.
- For a more detailed account, see Lead-Laced Water In Flint: A Step-By-Step Look At The Makings Of A Crisis (https://www.npr.org/sections/thetwo-way/2016/04/20/465545378/lead-laced-water-in-flint-a-step-by-step-look-at-the-makings-of-a-crisis).
- An excellent graph (that your students can replicate) appears in The Murky Tale of Flint's Deceptive Water Data by Robert Langkjaer-Bain (http://onlinelibrary.wiley.com/doi/10.1111/j.1740-9713.2017.01016.x/full).
- You might also want to read What Went Wrong In Flint by Anna Maria Barry-Jester (https://fivethirtyeight.com/features/what-went-wrong-in-flint-water-crisis-michigan/).

Potential Connections to Other Subjects

Depending on the level of students in your classroom, this activity has implications for collaborations with other disciplines. At the middle or high school level, a concurrent class in Social Studies or Chemistry could investigate other aspects of this issue. The authors are college professors and have used this activity in an undergraduate introductory Statistics course for General Education credit with students who are primarily majoring in disciplines other than Statistics. This allows for an extension where students investigate another angle of the story. Students majoring in Journalism, Criminal Justice, or Natural Resource Management will approach this from very different perspectives.