

How Wet is the Earth?

Laura Ring Kapitula
Grand Valley State University
kapitull@gvsu.edu

Paul Stephenson
Grand Valley State University
stephenp@gvsu.edu



Published: June 2014

Overview of Lesson

In this activity random sampling is used to estimate the proportion of the Earth's surface that is covered with water. Students use an internet site to select random points on the surface of the Earth and to see them on a map. After selecting their sample of points the students record whether or not each point is on water. Each student then uses their data to calculate the sample proportion of points that are on water and compute a confidence interval for the proportion of the Earth's surface that is covered by water.

After each student or student group finishes their calculations, the class' data can be used to illustrate the sampling distribution of the sample proportion and the long term behavior of confidence intervals. This lesson can also be adapted for use with middle school students if the confidence intervals are discussed only briefly or in a simplified manner.

GAISE Components

This activity follows all four components of statistical problem solving put forth in the *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report*. The four components are: formulate a question, design and implement a plan to collect data, analyze the data by measures and graphs, and interpret the results in the context of the original question. This is a GAISE Level C activity.

Common Core State Standards for Mathematical Practice

2. Reason abstractly and quantitatively.
4. Model with mathematics.
5. Use appropriate tools strategically.
8. Look for and express regularity in repeated reasoning

Common Core State Standards Grade Level Content (High School)

- S-IC. 1. Understand statistics as a process for making inferences about population parameters based on a random sample from that population.
- S-IC. 4. Use data from a sample survey to estimate a population mean or proportion; develop a margin of error through the use of simulation models for random sampling.

Common Core State Standards Grade Level Content (Grades 6 and 7)

6. SP. 1. Recognize a statistical question as one that anticipates variability in the data related to the question and accounts for it in the answers.
6. SP. 4. Display numerical data in plots on a number line, including dot plots, histograms, and box plots.

6. SP. 5. Summarize numerical data sets in relation to their context.
7. SP. 1. Understand that statistics can be used to gain information about a population by examining a sample of the population; generalizations about a population from a sample are valid only if the sample is representative of that population. Understand that random sampling tends to produce representative samples and support valid inferences.
7. SP. 2. Use data from a random sample to draw inferences about a population with an unknown characteristic of interest. Generate multiple samples (or simulated samples) of the same size to gauge the variation in estimates or predictions.
7. SP. 5. Understand that the probability of a chance event is a number between 0 and 1 that expresses the likelihood of the event occurring. Larger numbers indicate greater likelihood. A probability near 0 indicates an unlikely event, a probability around 1/2 indicates an event that is neither unlikely nor likely, and a probability near 1 indicates a likely event.
7. SP. 6. Approximate the probability of a chance event by collecting data on the chance process that produces it and observing its long-run relative frequency, and predict the approximate relative frequency given the probability.

**NCTM Principles and Standards for School Mathematics
Data Analysis and Probability Standards for Grades 9-12**

Formulate questions that can be addressed with data and collect, organize, and display relevant data to answer them:

- know the characteristics of well-designed studies, including the role of randomization in surveys and experiments;
- understand the meaning of measurement data and categorical data, of univariate and bivariate data, and of the term variable;
- understand histograms, parallel box plots, and scatterplots and use them to display data;
- compute basic statistics and understand the distinction between a statistic and a parameter.

Develop and evaluate inferences and predictions that are based on data:

- use simulations to explore the variability of sample statistics from a known population and to construct sampling distributions;
- understand how sample statistics reflect the values of population parameters and use sampling distributions as the basis for informal inference.

Understand and apply basic concepts of probability:

- use simulations to construct empirical probability distributions.

Prerequisites

Students should have some knowledge of sampling and estimating unknown parameters.

Learning Targets

Upon completion of the activity students will be able to:

- Randomly sample from a population.
- Estimate the proportion in a population with a certain characteristic.
- Understand variability in estimated proportions.
- Compute a confidence interval for a population proportion.

Time Required

This activity can be completed in one 50 minute class period.

Materials Required

Each student or student group needs to have a computer that is connected to the internet and that has a word processing program such as MS-Word. Having a printer in the classroom makes it a bit easier for students, but it is not absolutely necessary. Each student will need a copy of the worksheet that is given at the end of the lesson.

Alternatively, each student or student group needs to have an iPad or other tablet connected to the internet.

Instructional Lesson Plan

The GAISE Statistical Problem-Solving Procedure

I. Formulate Question(s)

What proportion of the Earth's surface is covered in water? The true proportion is what we are estimating, and we will use sample data to estimate this proportion. How will the sample proportions vary? What proportion of 90% confidence intervals calculated would we expect to contain the true proportion? Note that the true proportion is about 71%, see <http://ga.water.usgs.gov/edu/earthhowmuch.html> for more information.

II. Design and Implement a Plan to Collect the Data

Students should be instructed to go to <http://www.geomidpoint.com/random/> and select circular region and whole earth (see Figure 1). It may be a good idea to start by having students randomly generate a point on Earth and click to see it on a map. The map will come up centered on a blue pin. Students should be instructed to ignore the blue pin and look for the red pin. Then they should determine whether or not the red pin is on water (they may need to zoom in). [Note: This single point is generated to illustrate the process employed to collect data, and this point will not be used for any future calculations.] Next have students randomly generate 50 points on Earth and click to see them on a map (see Figure 2).

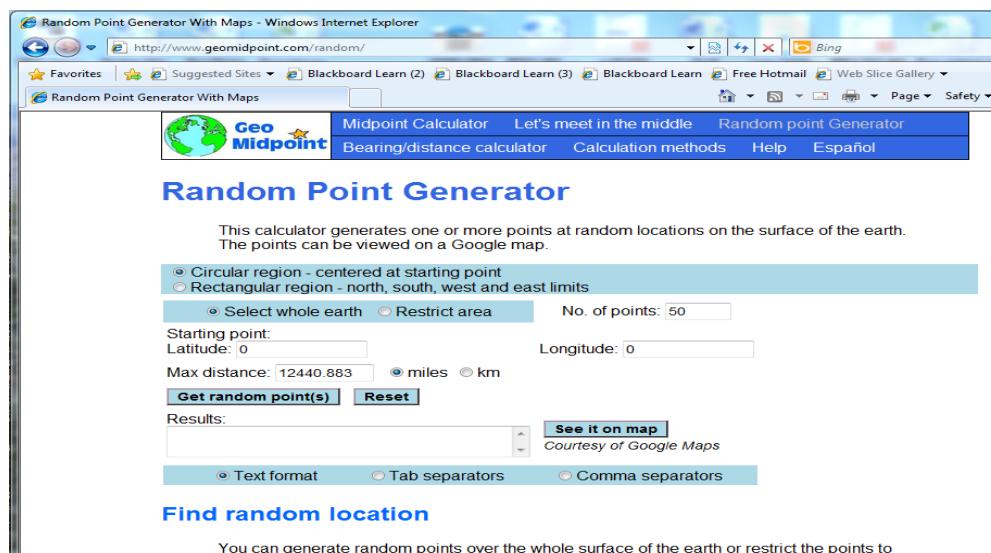


Figure 1. Generating 50 random points.

Remind students to be careful to not count points more than once. Given the rectangular nature of the map they may have to zoom in to make sure each pin is on the map only once. Students can then use the Snipping Tool to copy the map with the 50 points. For example:



Figure 2. Example of snipped image of map for printing.

After the map is generated and snipped, the students should open Word (or some other word processor) and paste the map into a Word file. They then can print the Word file and use the webpage and zooming to determine how many of the 50 points are on water, and circle the pins that are on the water. Once they do this they should record the number of pins on water and on land on their worksheet.

III. Analyze the Data

Students are then asked to pair up with a neighbor and combine results to create a sample size of 100. Using the combined data and the following formulas, students calculate a sample proportion of points that are on water and a 90% confidence interval for the proportion of the Earth that is covered with water:

$$n = \text{total number in sample}, \hat{p} = \frac{\text{number in sample on water}}{n}, \text{and the corresponding 90\% CI}$$

equals: $\hat{p} \pm 1.645 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, given the large sample method is appropriate in this case. Students could check the appropriateness of the large sample method by confirming that $n\hat{p} > 10$ and $n(1-\hat{p}) > 10$. It would be extremely unlikely to get a sample that would violate these conditions if a sample size of 100 is used, given the true proportion is around 71%.

Once students have an estimated proportion they can share their proportion with the class. A dot plot or stem- and- leaf plot with the sample proportions can be drawn on the board. Typically we allow students to come up to the board and write in their value on a plot rounded to the nearest hundredth. The plot of the sample proportions is an empirical sampling distribution for the sample proportion. See Figure 3 below.

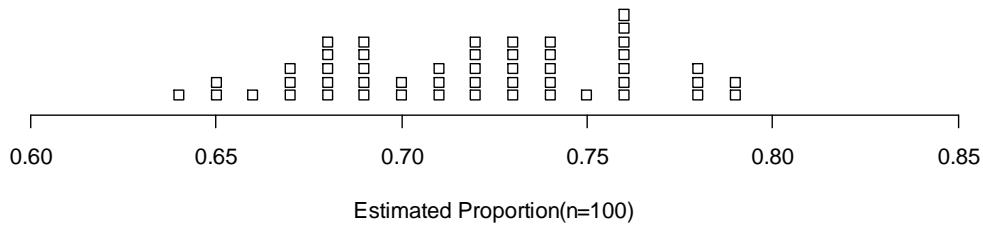


Figure 3. Example dot plot of student's estimated proportions for 50 samples of size 100.

After students generate their confidence intervals, ask the class groups whether or not their confidence interval contains the true value of 0.71. Tally the results and see what proportion of the 90% confidence intervals computed by the class contain the true value.

IV. Interpret the Results

Ask students how much spread they see in the empirical sampling distribution. Ask them why not everyone got the true value of 71% and ask them about the shape and center of the distribution of sample proportions. In the dot plot in Figure 3 the values are centered at about 71% and look fairly symmetric with no outliers. Ask them how they would expect the sampling distribution to change if instead of everyone in the class having a sample of 100, they had a sample size of 50, or a sample size of 1000. Simulated data can be used to illustrate this concept or the instructor can show the plot given in Figure 4. R-code to simulate data and make dot-plots and some simulated data are attached to this lesson.

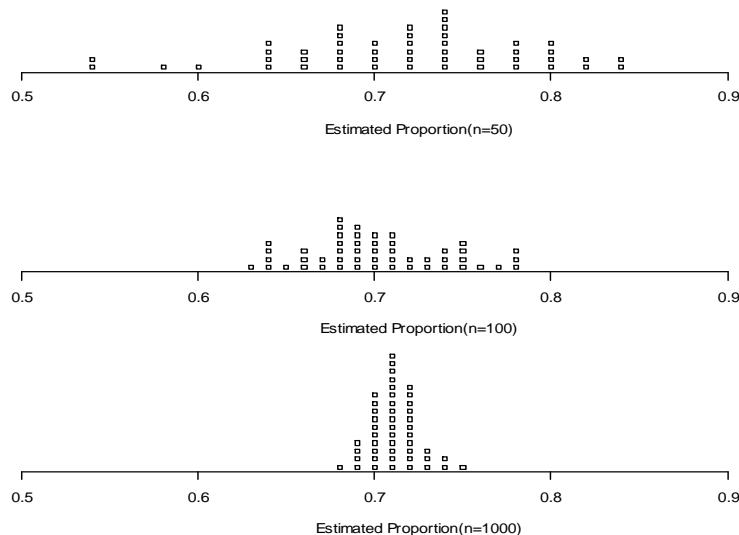


Figure 4. Simulated sampling distributions for the sample proportion.

If time is available the instructor can calculate a confidence interval using data from the whole class and talk about how the size of the margin of error is smaller when the sample size is larger.

Assessment

1. Suppose we did the activity again but had each student look at 500 points, instead of 50 points resulting in a sample size of 1000 when they combined their data with their partner. How would this change impact the distribution of the sample proportions? How would this change impact the 90% confidence interval?
 2. Suppose a student obtained 75 points out of 100 on the water. Use this data to calculate a sample proportion and both a 90% and a 95% confidence interval for the proportion of water on the surface of the Earth. Which interval is wider? Will this always be the case?

Answers

1. Increasing the sample size would result in the sample proportions being less variable. They would be in general closer to 71% but still centered at 71%. They would be more bell shaped as well but students may not get that from this lesson since given the true population proportion is moderate the sampling distribution should be fairly bell shaped with an $n=100$. Increasing the sample size would cause the confidence intervals to be less wide, but they still would have a success rate of 90% since the success rate is fixed by the method.

2. $\hat{p} = 0.75$, the number of failures is 25 and the number of successes 75 so it is reasonable to use the large sample method to calculate the confidence interval for the proportion.

The 90% confidence interval is $\hat{p} \pm 1.645\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = (0.679, 0.821)$. And the 95% confidence interval is $\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = (0.665, 0.835)$. Students should notice that the 95% confidence interval is wider than the 90% confidence interval and looking at the formulas they should notice that this will always be true.

Possible Extensions

- What proportion of the surface area of the southern or northern hemisphere is covered with water?
- Is the proportion of water in the southern and northern hemisphere different?
- The STEW activity: “What Percent of the Continental US is Within One Mile of a Road?” by Stoudt, Cao, Udwin and Horton is an excellent extension or alternative to this activity for students who are using R in class. We developed this activity independently before reading Stoudt, et. al., but there are many similarities with the Stoudt, et. al. activity.
- This activity could tie in with learning to calculate the surface area of a sphere and using that value to estimate the surface area of land mass on the Earth’s surface.

Possible Simplifications

- Do not estimate confidence intervals and only use the activity to illustrate the sampling distribution for a proportion.
- For elementary or middle school students they could get in groups and look at 100 points and use the information to calculate and understand the idea of a percentage.
- Younger children could sample points and learn to make a pie chart or bar chart using their sample or a larger sample for the whole class.

Acknowledgements

We wish to thank Randall Pruim who inspired this lesson and implemented the "rgeo()" and "googleMap()" functions in the mosaic package for the software package R. Although we did not use R in designing our lesson, his work and a personal discussion inspired the activity. If you wish to implement this activity in R, modifications can be made to the STEW activity: “What Percent of the Continental US is Within One Mile of a Road?” by Stoudt, Cao, Udwin and Horton.

References

1. *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report*, ASA, Franklin et al., ASA, 2007. <http://www.amstat.org/education/gaise/>
2. *Common Core State Standards for Mathematics*, Common Core State Standards Initiative (CCSSI). Released June 2, 2010. <http://www.corestandards.org/>
3. *Principles and Standards for School Mathematics*. National Council of Teachers of Mathematics (NCTM). Reston, VA: NCTM, 2000.
4. What Percent of the Continental US is Within One Mile of a Road? Stoudt, Cao, Udwin and Horton, ASA, 2014. <http://www.amstat.org/education/stew/>
5. Water drop image taken from: http://naturesan.com/images/earth_water_drop.png

How Wet is the Earth?



Activity Sheet

We will take random samples of points on the Earth's surface and see if we can use these samples to learn about estimating the proportion of the Earth's surface that is water.

1. What numeric value (or parameter) are we estimating?

2. Go to <http://www.geomidpoint.com/random/> and select circular region and whole Earth.

3. Now randomly generate a point on Earth and click to see it on a map. The map will come up centered on a blue pin. Ignore that blue pin and look for the red pin. Determine whether or not the pin is on water (you may zoom in if you need to).

4. Now randomly generate 50 points on Earth and click to see the points on a map.
 - Use the Snipping Tool to copy the map with the 50 points.
 - Open Word and paste the map into a Word file.
 - Print the Word file with the map.
 - Go back to the webpage, and determine how many of the 50 points are on water, zoom in if you need to.
 - Circle pins on the map that are in the water.

5. Record the number of pins on water and on land.

On Water	On Land

6. Pair up with one of your neighbors.
7. Combine the results with your neighbor.



On Water	On Land

8. Calculate the sample proportion of points on water for these combined samples.
9. Compute a 90% confidence interval for the proportion of points on water. This is an estimate for the proportion of the Earth covered by water.

As you may or may not know about 71% of the Earth's surface is covered by water. See <http://ga.water.usgs.gov/edu/earthhowmuch.html> for more information.