

# It Creeps. It Crawls. Watch Out For *The Blob*!

John Gabrosek  
Grand Valley State University  
[gabrosej@gvsu.edu](mailto:gabrosej@gvsu.edu)



**Published: September 2014**

## Overview of Lesson

This lesson uses data collected by students to estimate the area of an irregularly shaped three-dimensional object using sampling, sampling variability of the sample mean, and confidence interval estimation. The approach of estimating area using a confidence interval is an alternative to the Calculus approach of integrating to find the area. The advantage of using a confidence interval is that one does not need to know the actual function,  $f(x)$ , describing the object's shape.

The irregularly shaped object under study is *The Blob*. *The Blob* is the name and lead character of an independently made movie released by Paramount Pictures in 1958. A remake was released in 1988. The 1958 movie starred a young Steve McQueen who was 27 years old and played a rambunctious teenager. Ah Hollywood.

Because *The Blob* is a three-dimensional object, when we discuss finding its area, we actually are finding the area of a cross-section of *The Blob*.

The activity uses a gridded scale drawing of *The Blob*. Students generate random numbers, determine lengths on a grid, multiply by a scaling factor, and find area estimates for squares. These area estimates are the data. Students summarize these data numerically and graphically with tools such as histogram, boxplot, five-number summary (minimum, first quartile – i.e. the 25<sup>th</sup> percentile, median, third quartile – i.e. the 75<sup>th</sup> percentile, maximum), mean, and standard deviation. Each student creates her own confidence interval estimate of the area of a cross-section of *The Blob*. The different confidence interval estimates of the students are compared, which leads to discussion of sampling variability and the confidence level.

## GAISE Components

This investigation follows the 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

1. Make sense of problems and persevere in solving them.
2. Reason abstractly and quantitatively.
4. Model with mathematics.
5. Use appropriate tools strategically.
6. Attend to precision.

### **Common Core State Standards Grade Level Content (High School)**

S-ID. 1. Represent data with plots on the real number line (dot plots, histograms, and box plots).

S-ID. 3. Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).

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.

### **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 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.

**Select and use appropriate statistical methods to analyze data:**

- for univariate measurement data, be able to display the distribution, describe its shape, and select and calculate summary statistics.

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

- understand how sample statistics reflect the values of population parameters and use sampling distributions as the basis for informal inference.

### **Prerequisites**

Students will have knowledge of calculating numerical summaries for one variable (mean, median, five-number summary, checking for outliers). Students will have knowledge of how to construct appropriate graphical summaries for univariate quantitative data (histograms and boxplots). Students will be familiar with the mechanics of the calculation of a t-distribution confidence interval on the population mean.

### **Learning Targets**

Students will be able to calculate numerical and graphical summaries and use them to describe a data set. Students will be able to construct a confidence interval on the population mean.

Students will be able to explain the impact of an outlying data value on sample-based summaries such as the mean, median, standard deviation and confidence interval.

### **Time Required**

1 to 2 class periods. The activity description and data collection will take approximately 30 minutes. Allow approximately 45 minutes for answering questions and class discussion.

### **Materials Required**

Pencil and paper; graphing calculator; statistical software package (optional – used to automate calculations); gridded figure of *The Blob* (see page 18).

## Instructional Lesson Plan

### The GAISE Statistical Problem-Solving Procedure

#### I. Formulate Question(s)

Ask students how they could find the area of an irregularly shaped object. You may want to show a picture of something like a golf green, Figure 1, or a diatom, Figure 2, to motivate that this is a real problem of interest.



Figure 1. Pebble Beach golf course.



Figure 2. An amoeba.

Next, introduce the gridded scale drawing picture of *The Blob* on page 18. You may also want to show a picture of the original movie poster from the 1958 film (Figure 3).



Figure 3. Movie poster for the 1958 film.

Explain to students that because *The Blob* is a three-dimensional object, when we speak of finding its area we actually are focusing on the area of a cross-section of *The Blob*. You may choose to describe this as the foot-print of *The Blob*.

Discuss with students that they should be able to find a lower estimate and an upper estimate for the area of the cross-section of *The Blob* depicted in the gridded scale drawing on page 18. Ask students to find lower and upper estimates and write a short description in complete sentences of the process used (Questions 1 and 2 on the Activity Sheet, pages 13 and 14.) These questions have students practice estimation skills. Of course, the advantage to using a confidence interval as an estimation technique is that personal selection bias is not present. In other words, how adept the student is at choosing lower and upper estimates is not a factor.

Discuss with students that if we were to use Calculus techniques, we could only find the exact area of a cross-section of *The Blob* if we had a functional description of the shape. Point out to students that for irregularly shaped objects, knowing the functional description of the object's shape is usually unrealistic. Explain that statistical inferential techniques are just one tool among many for estimation.

## II. Design and Implement a Plan to Collect the Data

The data collection plan is pre-determined. Students must carry out the data collection plan by generating random numbers and finding lengths on a gridded scale drawing. Students complete a Data Table as they proceed. The Data Table is contained in the Activity Sheet (See page 14). Be sure that students realize that they are generating random numbers between 1 and 70 and that they need to locate the correct random number grid line on the horizontal axis of the gridded scale drawing. At this location, they will determine the vertical grid distance (Blob Height) measurement. Students will need to generate twelve random integers.

To generate random numbers using the TI-calculator complete the following:

**Enter a Seed Value in the TI-83/84**

Have each student choose any positive whole number for seed value, such as the last four digits of the student's phone number. For example, to enter a seed value of 4092,

4092 STO MATH PRB 1:RAND ENTER ENTER

The calculator will show your seed value. You only need to do this one time for your calculator.

**Generating Random Integers on the TI-83/84**

To generate a list of random integers between 1 and 70 press:

**MATH PRB 5:RANDINT(1,70) ENTER**

Continue to push ENTER. Your calculator will generate a list of random numbers between 1 and 70.

We have chosen to use a random number generator to take a simple random sample of the grid locations. Other sampling designs are possible. For example, one could choose to use evenly spaced intervals to find a sample of grid locations. In essence, this would be a systematic 1-in- $k$  sample where every  $k^{\text{th}}$  grid location is sampled. We choose to use a random number generator to take a simple random sample for a couple of reasons. First, while there is no guarantee that any one particular simple random sample will result in a sample that is representative of the population, other sampling techniques are often more likely to result in non-representative samples. In particular, the representativeness of a systematic sample is very dependent on a fortuitous choice of  $k$  and the first value sampled (usually called the starting point) in a 1-in- $k$  sampling design. If an unlucky choice of  $k$  or starting point is made it can occur that the sample follows a periodicity in the population and thus is not representative of the population as a whole. For example, suppose the shape of *The Blob* was very predictably periodic, such as if the functional form of *The Blob* followed a sine curve. If  $k$  was chosen to match this periodicity, the sample would either consistently choose grid locations that either overestimate or underestimate *The Blob's* area. Second, the confidence interval techniques typically taught in introductory statistics courses use formulas that are valid for simple random samples. That is, the formula is based on understanding the sampling variability in the sample mean generated by taking simple random samples from the population.

Ask students: What geometric shape are we using to estimate the area of a cross-section of *The Blob*? You might want to discuss how this may introduce error into the estimation process. It's important for students to realize that the technique works relatively well because the shape being estimated is somewhat square and that underestimates and overestimates of the area will balance out when we combine multiple estimates to find the sample mean and make the confidence interval.



### III. Analyze the Data

Begin data analysis by asking students to suggest graphs that might be used to summarize the data. Dotplot, boxplot, or histogram are appropriate tools. Then ask students to find the five-number summary (minimum, first quartile – i.e. the 25<sup>th</sup> percentile, median, third quartile – i.e. the 75<sup>th</sup> percentile, maximum), mean, and standard deviation of their twelve values. Technology may be used to make the calculations easier. If you want students to actually construct a plot, the boxplot is easiest since students will have the five-number summary.

The remainder of the lesson is based on students constructing a confidence interval on the population mean. The population mean is really an estimate of the area of a cross-section of *The Blob*. Students should already be familiar with the mechanics of constructing a confidence interval. If they are not, discussion time must be included to provide students with an illustrative example.

Have students make a confidence interval on the mean. This can be done by hand, using technology, or both. The proper formula to use is:  $\bar{x} \pm t^* \left( \frac{s}{\sqrt{n}} \right)$ . Because each student has

$n = 12$  measurements of *The Blob*'s area and each student is making a 90% confidence interval, the confidence level multiplier is:  $t^* = 1.796$ . This can be given to students or found using a table. Many textbooks include such a table. For the TI-84 calculator use 2<sup>nd</sup> – Distr –

4:invT(.05,11), where .05 comes from  $1 - \frac{\alpha}{2}$ ,  $\alpha = .10$ , and the degrees of freedom are  $n - 1 = 11$ .

Online tools are readily available as well. (For example, Webster West's applet can be found at <http://www4.stat.ncsu.edu/~west/applets/tdemo.html>, Java must be enabled). Once the interval is made, students should write a sentence or two interpreting the interval in the context of the problem.

Ask students: Suppose you had an area estimate for *The Blob* that is much smaller than all the other area estimates. What impact would you anticipate this outlying value having on the mean, the median, the standard deviation, and the 90% confidence interval? Students should already be familiar with the impact of outliers on the mean, median, and standard deviation. It may take them some time to realize that a single outlying value on the low end will move the center of the confidence interval toward it and increase the margin of error because of the increase in the standard deviation. This leads naturally to a discussion of: Why is making a confidence interval based on 12 estimates of the area of a cross-section of *The Blob* superior to just generating one random number and then finding a single estimate of the area? A discussion of the tradeoffs in using larger sample sizes such as 30 or 40 and the associated impacts on the confidence interval is appropriate. As the sample size  $n$  increases, the sampling variability in the sample mean, as estimated by  $\frac{s}{\sqrt{n}}$ , decreases and the confidence interval becomes narrower. A narrower

confidence interval is a more precise estimate of the area of *The Blob*. The tradeoff is that in many situations collecting data requires a substantial investment in time and resources. Thus, an increase in the sample size requires more time and resources to collect the data.

Have students write their confidence interval values on the whiteboard/chalkboard so that all confidence intervals for the class can be viewed.

#### IV. Interpret the Results

Ask students: Why did different students arrive at different confidence interval values? Students should relate the differences back to the differences in the sample mean and sample standard deviation. This can then be related back to the fact that different students had different random samples (i.e. different random numbers). Discuss with students that a sample returns only one confidence interval but a sample is only one sample from among all the possible samples drawn from the population. Thus, the one confidence interval calculated for a particular sample is only one confidence interval from among all the possible confidence intervals estimating the population mean. This provides a very nice bridge from the sampling distribution on the sample mean to the confidence interval on the population mean.

Figure 4 shows the relationship between sampling, the sampling distribution of the sample mean, and the confidence interval on the population mean. One thousand separate simple random samples of size 12 were taken from the population of grid locations depicted in the scale drawing of *The Blob* on page 17. Each of these samples generates a sample mean. In Figure 4, these 1000 sample means are shown in Graph A, which is a graph of the sampling distribution of the sample mean. Different random samples lead to different sample averages, which vary from 10,000 to 20,000 square feet.

Each sample can be used to construct a confidence interval on the population mean. In Figure 4 Graph B, the confidence intervals constructed from the first 25 random samples are graphed. Different random samples lead to different confidence intervals. Most of the confidence intervals include the actual area of *The Blob*, which is about 15,095 yds<sup>2</sup>, within the interval. However, the occasional interval does not. In Graph B, a vertical line is drawn at 15,095. Of the 25 intervals shown, there are 24 that include this value within the interval. Only the 8<sup>th</sup> interval from the top fails to include this value.

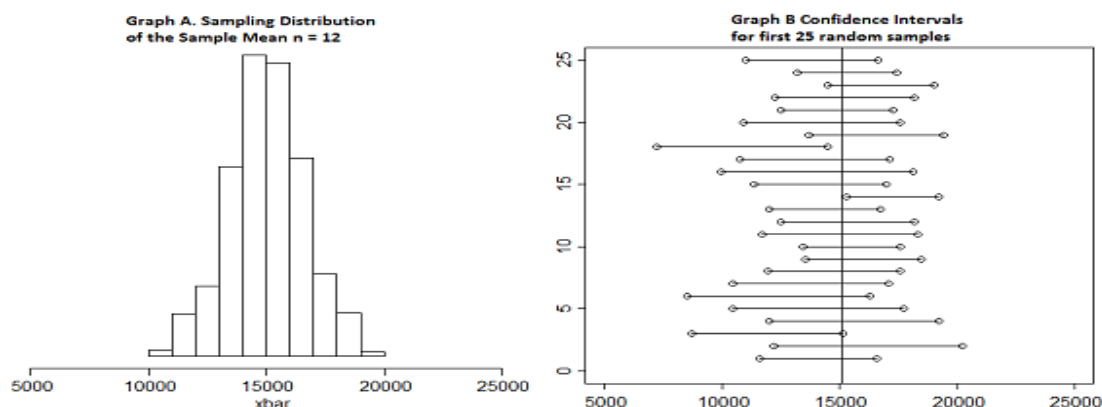


Figure 4. Simulation of the sampling distribution of the sample mean.

Tell students: The area of the cross-section of *The Blob* is thought to be about 15,000 yds<sup>2</sup>.

Ask students: Is 15,000 inside your confidence interval? How many of the class' 90% confidence intervals include the value 15,000? What percent is this? Suppose that instead of having only our class' confidence intervals, we had confidence intervals created from many classes – 1000 confidence interval estimates of *The Blob*'s area! What percentage of these interval estimates do we anticipate would contain 15,000? Write a sentence or two that explains what we mean by the confidence level (i.e. the 90%) in this problem. Think about this in the context of repeatedly taking random samples of the same size from the same population.

The goal of these questions and this discussion is to have students begin to understand the concept of confidence level. Previous questions will help students to see that different random samples lead to different confidence intervals. This series of questions helps students to see that occasionally a random sample will produce a confidence interval that does not contain the population parameter being estimated. Of course, in reality we don't know the actual area of a cross-section of *The Blob*. If we did know the "true value" of the population mean there would be no need to estimate it with a confidence interval. While we cannot know if a particular confidence interval contains the population mean, we can discuss what would happen if we knew the population mean. The confidence level tells us how often a random sample would return a confidence interval containing the value of the population mean.



### Assessment

1. For each of the following state whether *an increase in the value* will increase, decrease, or not change the margin of error of a confidence interval on the population mean assuming that all other quantities in the confidence interval formula stay the same.

(a) The sample mean

(b) The sample standard deviation

(c) The confidence level

(d) The sample size

(e) The population size

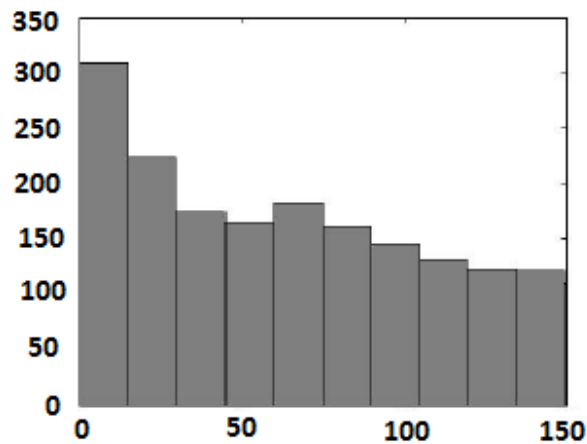
2. Suppose that we take a random sample of 100 University of Michigan football fans who are attending a fall game in “The Big House.” A 95% confidence interval for the mean amount spent per person on food is found to be (\$4.62, \$7.74). Decide whether or not each statement given below is a correct conclusion that can be drawn from the confidence interval.

(a) We can be pretty sure that the average amount spent on food per person for all the fans at the game is between \$4.62 and \$7.74.

(b) About 95% of the people at the game spent between \$4.62 and \$7.74 on food.

(c) The average amount spent on food per person for the 100 sampled fans is \$6.18.

3. The paper “Self-exciting point process modeling of crime” by Mohler et al. (2011) contains a histogram similar to that given below and shows the time between burglary events separated in space by 180 meters or less, for 2004 residential burglaries within an 18km by 18km region in the Los Angeles area. The x-axis is measured in days. The vertical axis is the number of burglary events. Total number of burglary events is about 1950. The paper does not report the mean or standard deviation of the days between events but approximate values found from the graph are: mean = 63 days and standard deviation = 45 days.



(a) Make a 95% confidence interval for the mean time between burglary events.

(b) Interpret your interval in the context of the problem.

## Answers

1. (a) Not change. The sample mean locates the center of the confidence interval. The value of the sample mean has no effect on the variability in sample means from random sample to random sample. Thus, no impact on the margin of error.

(b) Increase. The standard deviation is a measure of variability. An increase in the standard deviation equates to an increase in variability in the individual data values. This leads to an increase in the variability in sample means from random sample to random sample. Thus, an increase in the margin of error.

(c) Increase. The confidence level is a measure of how likely random sampling is to produce a confidence interval that contains the population parameter. An increase in the confidence level means that you want greater confidence that the procedure being used will produce an interval that contains the population parameter. To increase this probability, more possible values for the population mean need to be included inside the interval. Thus, an increase in the margin of error. In the t-interval formula, increasing the confidence level increases the confidence level multiplier  $t^*$ .

(d) Decrease. Increasing the sample size means that you have more information on the population. The more information that you have the more accurate (i.e. closer to the true value) you expect the sample mean to be. Hence, a decrease in the margin of error. In the t-interval formula, increasing the sample size implies that you are dividing by a larger value of  $n$ .

(e) No change. Provided the sample size is small relative to the population size, increasing the population size has no effect on the margin of error. Nor does increasing the population size have any effect on how big a sample you need to take. The key is not size of population in random sampling. The key is that the random sample needs to be representative of the population. If you can represent a population with a sample of size 100, it doesn't matter if the population has size 100 thousand, 100 million, or 100 billion!

2. (a) Correct. A confidence interval allows us to use the sample data to draw conclusions about the mean of a population. In this situation we can conclude about the mean amount spent on food per fan for the population of all people at the game.

(b) Incorrect. A confidence interval allows us to draw conclusions about the mean, not individual values. Individual values are much more variable than is the mean of 100 values. Far fewer than 95% of the individual fans will spend between \$4.62 and \$7.74 on food.

(c) Correct. The formula used to compute the confidence interval guarantees that the sample mean will always be at the center of the interval. The center of this interval is \$6.18. Note: The confidence interval does not tell us that the population mean is in the middle of the interval.

3. (a) 
$$\bar{x} \pm t^* \left( \frac{s}{\sqrt{n}} \right) = 63 \pm 1.961 \left( \frac{45}{\sqrt{1950}} \right) = 63 \pm 2 = (61, 65) \text{ days}$$

(b) We are 95% confident that the mean time to a second burglary event in Los Angeles within the 18km by 18km region is between 61 and 65 days. It is important that students are specific in defining that inference is restricted to the 18km by 18km region of Los Angeles on which the data were collected.

## Acknowledgement

The author thanks the Editor, Associate Editor, and Referee for providing valuable recommendations that improved the lesson. This lesson is based on an activity that was published by the author in a SumMore-Math workbook (Gabrosek 2012).

## References

1. *Guidelines for Assessment and Instruction in Statistics Education* (GAISE) Report, ASA, Franklin et al., ASA, 2007 <http://www.amstat.org/education/gaise/> Accessed August 15, 2012
2. *Common Core State Standards for Mathematics*, [http://www.corestandards.org/assets/CCSSI\\_Math%20Standards.pdf](http://www.corestandards.org/assets/CCSSI_Math%20Standards.pdf) Accessed August 15, 2012
3. NCTM Principles and Standards for School Mathematics: Data Analysis and Probability Standards for Grades 9-12, <http://www.nctm.org/standards/content.aspx?id=318> Accessed August 15, 2012
4. Pebble Beach photo, [https://encrypted-tbn1.google.com/images?q=tbn:ANd9GcQdzRTiSdo\\_yGOkJEKcIirWm8euhTHVYVExKWx-1sAZitIYLdgK](https://encrypted-tbn1.google.com/images?q=tbn:ANd9GcQdzRTiSdo_yGOkJEKcIirWm8euhTHVYVExKWx-1sAZitIYLdgK) Accessed August 15, 2012
5. Amoeba photo, <https://encrypted-tbn2.google.com/images?q=tbn:ANd9GcSG9cAhQFYE0bpBBukS73hQrkbwg6EWoFxmj7Ee7fSluf93Mnvqag> Accessed August 15, 2012
6. The Blob photo, [http://www.google.com/imgres?um=1&hl=en&client=firefox-a&sa=N&rls=org.mozilla:en-US:official&biw=1366&bih=622&tbn=isch&tbnid=U-u91pi00yGkqM:&imgrefurl=http://www.asharperfocuse.com/theblob.htm&docid=frnig5-74E\\_izM&imgurl=http://www.asharperfocuse.com/THE\\_BLOB-4.jpg&w=852&h=480&ei=mu0rULSvO-GUywGy6IDIBw&zoom=1&iact=rc&dur=405&sig=109753512354896112459&page=1&tbnh=92&tbnw=163&start=0&ndsp=23&ved=1t:429,r:7,s:0,i:163&tx=98&ty=17](http://www.google.com/imgres?um=1&hl=en&client=firefox-a&sa=N&rls=org.mozilla:en-US:official&biw=1366&bih=622&tbn=isch&tbnid=U-u91pi00yGkqM:&imgrefurl=http://www.asharperfocuse.com/theblob.htm&docid=frnig5-74E_izM&imgurl=http://www.asharperfocuse.com/THE_BLOB-4.jpg&w=852&h=480&ei=mu0rULSvO-GUywGy6IDIBw&zoom=1&iact=rc&dur=405&sig=109753512354896112459&page=1&tbnh=92&tbnw=163&start=0&ndsp=23&ved=1t:429,r:7,s:0,i:163&tx=98&ty=17) Accessed August 15, 2012
7. West, W. t-distribution applet <http://www.stat.tamu.edu/~west/applets/tdemo.html> Accessed August 15, 2012
8. Mohler, G.O., Short, M.B., Brantingham, P.J., Schoenberg, F.P., and Tita, G.E. (2011). "Self-Exciting Point Process Modeling of Crime", *Journal of the American Statistical Association*, **106**, 100-108.
9. Gabrosek, J. (2012). "The Blob," *SumMore-Math Adventures in Mathematics: Statistics and Probability*, Charlene Beckmann and Mary Richardson (eds.), Michigan Council of Teachers of Mathematics.

## **The Blob Activity Sheet**

**Name:** \_\_\_\_\_

This lesson is motivated by the classic horror film, “*The Blob*” (1958). The following information on the movie is taken from the Internet Movie Database (<http://www.imdb.com/title/tt0051418/synopsis>).

*July 20, 1957. In the small Pennsylvania town of Downington, teenager Steve Andrews (Steve McQueen) and his girlfriend Jane Martin (Aneta Corsaut) are out parking and see a falling star. They drive out to try to find where the meteor landed. An old man (Olin Howland) has heard the meteor crash near his house. He finds the meteor and pokes it with a stick. The rock breaks open to reveal a small jelly-like blob inside. This Blob, a living creature, crawls up the stick and attaches itself to his hand. Unable to scrape or shake it loose (and apparently now in pain), the old man runs hysterically onto the road, where he is seen by Steve, who takes him to see the local doctor, Doctor Hallen.*

*They reach the clinic when Doc Hallen is about to leave. Hallen anesthetizes the man and sends Steve back to the crash site to gather more information. Hallen decides he must amputate the man's arm which is being consumed by the Blob, calling in his nurse. However, the Blob completely consumes the old man. Now an amorphous creature, it eats the nurse and the doctor while increasing in size.*

*Steve and Jane return to the office and Steve witnesses the doctor's death. They go to the local police and return to the clinic with the kindly Lt. Dave (Earl Rowe) and cynical Sgt. Bert (John Benson). However, there is no sign of the creature or the doctor, and the police dismiss Steve's story. Steve and Jane are sent home with their fathers but sneak out and retrieve Steve's friends and successfully enlist their help warning the town...*

We may not be able to stop *The Blob*, pretty much nobody can, but we can figure out how big it is. Attached on a separate sheet of paper is a Gridded Scale Drawing of *The Blob*. Each 1 grid square on the drawing is equivalent to 2.2 yards. Our goal is to estimate the size of *The Blob*. Notice from the Gridded Scale Drawing that *The Blob* has width and height that are roughly equal.

**1.** You should be able to find a lower estimate and an upper estimate for the area of the cross-section of *The Blob* depicted in the gridded scale drawing. Find lower and upper estimates.

Lower estimate = \_\_\_\_\_ Upper Estimate = \_\_\_\_\_



2. Write a short description in complete sentences of what you did to find the lower and upper estimates in question 1.

### Data Collection

Follow the procedure below to collect data on the estimated size of *The Blob*.

- Generate a random real number from 1 to 70. Fill in the column “Random Number” in the Data Table.
- On the horizontal (x) axis in the Gridded Scale Drawing go to the grid line corresponding to the random number generated.
- Measure the height of *The Blob* by counting the number of vertical grid squares taken up by *The Blob* at the random number location on the horizontal axis. *Do not* include portions where there is no *Blob* present! Fill in the column Vertical Grid Squares.
- Fill in the column “*The Blob* Height” in the Data Table. Be sure to multiply the number of Vertical Grid Squares by 2.2.
- Fill in the column “*The Blob* Area” in the Data Table using the formula:  
 $Area = height \times height$ .
- Repeat the above process until you have 12 estimates for *The Blob*’s area.

Table 1: *The Blob* Data

Random Number	Vertical Grid Squares	<i>The Blob</i> Height (recall 1 grid square equals 2.2 yards)	<i>The Blob</i> Area

3. What geometric shape are we using to estimate the area of a cross-section of *The Blob*?

4. Suggest a graph or two that might be used to look at the area estimates. What would you be looking for in these graphs?

5. Calculate the five-number summary, mean, and standard deviation of *The Blob* area estimates. Then, determine if there are any outlying values.

Five-number summary = \_\_\_\_\_

Mean = \_\_\_\_\_ Standard Deviation = \_\_\_\_\_

6. Find a 90% confidence interval for the area of a cross-section of *The Blob*. Note that *The Blob* area estimates in the population are roughly normally distributed.

Lower confidence limit = \_\_\_\_\_ Upper confidence limit = \_\_\_\_\_

7. Interpret the meaning of your confidence interval in the context of the problem.

8. Suppose you had an area estimate for *The Blob* that is much smaller than all the other area estimates. What impact would you anticipate this outlying value having on:

a. the mean

b. the median

c. the standard deviation

d. the 90% confidence interval

9. Why is making a confidence interval based on 12 estimates of the area of a cross-section of *The Blob* superior to just generating one random number and then finding a single estimate of the area?

**Add your Confidence Interval estimate to the picture on the whiteboard/chalkboard. Wait until all the intervals are added to answer the remaining questions.**

10. Why is it that different students arrive at different confidence interval values?

**11.** The true area of a cross-section of *The Blob* is thought to be about 15,000 yds<sup>2</sup>. For reference, the area of a football field is 5300 yds<sup>2</sup>.

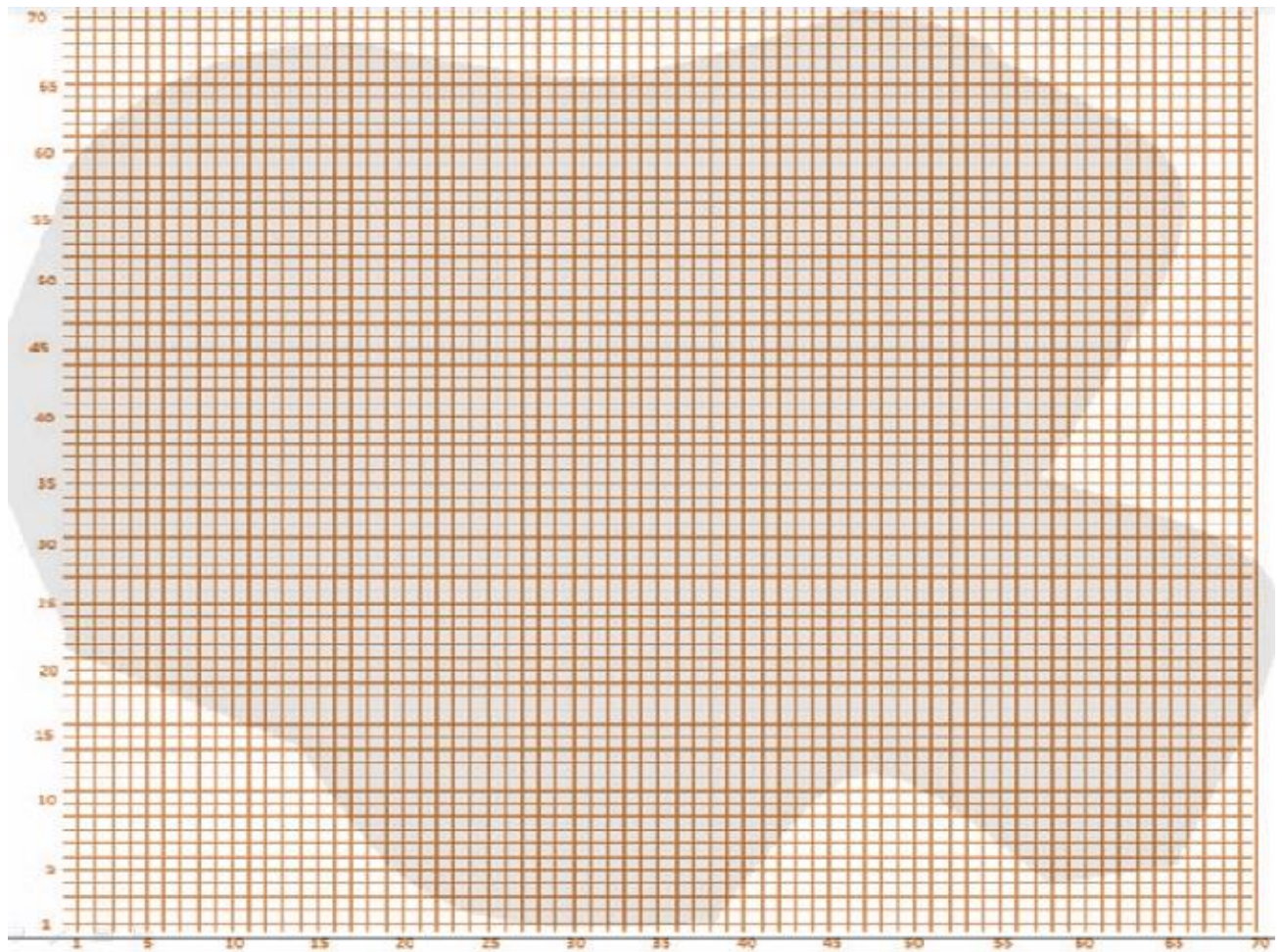
**a.** Is 15,000 inside your confidence interval?

**b.** How many of the class' 90% confidence intervals include the value 15,000? What percent is this?

**c.** Suppose that instead of having only your class' confidence intervals we had confidence intervals created from many classes – 1000 confidence interval estimates of *The Blob*'s area! What percentage of these interval estimates do we anticipate would contain 15,000?

**12.** Write a sentence or two that explains what we mean by the confidence level (i.e. the 90%) in this problem. Think about this in the context of repeatedly taking random samples of the same size from the same population.

***The Blob – Gridded Scale Drawing where 1 grid square = 2.2 yards***





***The Blob – Grid Height, Blob Height, and Blob Area for Random Numbers 1 through 70***

<b>Random Number</b>	<b>Grid</b>	<b>Blob</b>	<b>Blob Area</b>
1	38	83.6	6988.96
2	39	85.8	7361.64
3	42	92.4	8537.76
4	44	96.8	9370.24
5	45	99	9801
6	45	99	9801
7	45	99	9801
8	46	101.2	10241.44
9	46	101.2	10241.44
10	48	105.6	11151.36
11	51	112.2	12588.84
12	51	112.2	12588.84
13	52	114.4	13087.36
14	53	116.6	13595.56
15	57	125.4	15725.16
16	58	127.6	16281.76
17	59	129.8	16848.04
18	60	132	17424
19	62	136.4	18604.96
20	62	136.4	18604.96
21	63	138.6	19209.96
22	64	140.8	19824.64
23	65	143	20449
24	65	143	20449
25	65	143	20449
26	66	145.2	21083.04
27	65	143	20449
28	65	143	20449
29	65	143	20449
30	65	143	20449
31	65	143	20449
32	65	143	20449
33	65	143	20449
34	65	143	20449
35	66	145.2	21083.04
36	67	147.4	21726.76
37	67	147.4	21726.76
38	66	145.2	21083.04
39	66	145.2	21083.04
40	64	140.8	19824.64
41	63	138.6	19209.96
42	61.5	135.3	18306.09
43	61	134.2	18009.64
44	60	132	17424
45	59	129.8	16848.04
46	58	127.6	16281.76
47	58	127.6	16281.76
48	58	127.6	16281.76
49	58	127.6	16281.76

50	59	129.8	16848.04
51	58	127.6	16281.76
52	59	129.8	16848.04
53	54	118.8	14113.44
54	50	110	12100
55	60	132	17424
56	61	134.2	18009.64
57	62	136.4	18604.96
58	58	127.6	16281.76
59	62	136.4	18604.96
60	55	121	14641
61	52	114.4	13087.36
62	50	110	12100
63	47	103.4	10691.56
64	42	92.4	8537.76
65	36	79.2	6272.64
66	29	63.8	4070.44
67	24	52.8	2787.84
68	21	46.2	2134.44
69	16	35.2	1239.04
70	13	28.6	817.96