PROJECT-SET
Statistics Education for Teachers

How Many Have School Spirit?

Overview of Lesson

This activity poses the question "If all students at your high school were asked to wear the school color today, what is the best estimate of the amount of students who did?" Students are asked to explore random samples, make a dot plot using the results from the random sampling, and draw inferential conclusions from the sample statistic about the population parameter.

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 B and C lesson.

Common Core State Standards for Mathematical Practice

1. Reason abstractly and quantitatively.
2. Model with mathematics.
3. Use appropriate tools strategically.

Common Core State Standards Grade Level Content (Middle School & High School)

6.SP. 1. Develop understanding of sampling variability

7.SP. 1. Use random sampling to draw inferences about a population

7.SP. 3. Investigate chance processes and develop, use, and evaluate probability models

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

S-IC.1. Understand statistics as a process for making inferences about population parameters based on a random sample from that population.

S-IC.3. Recognize the purposes of and differences among sample surveys, experiments, and observational studies; explain how randomization relates to each.

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.

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

Prior to this activity, students must have knowledge of dot plots, random samples, and calculating the mean.

Learning Targets

In this lesson, students will be able to explore a population parameter, use repeated samples of the same size, clarify that the summary statistic varies in a predictable way, define and construct an approximation to the sampling distribution, and draw inferential conclusions from the sample statistic about the population parameter.

Time Required

The estimated time required for this activity is 45 minutes. This time will vary according to the method used to collect data.

Materials Required

For this activity, students will need a pencil, paper, and a calculator/or software. The instructor will provide the activity sheet. The data set School Data is also needed.

Instructional Lesson Plan
The GAISE Statistical Problem-Solving Procedure
I. Formulate the Question: For a given population parameter, how can we tell which summary statistics from the given population are common and which are unusual?

II. Collect Data: Initially, students are given a scenario where a random sample was selected in their school pertaining to students wearing the school color to school that day. They are then given a situation where 7 other students independently conducted the same random sample. The results are given for students to discuss and utilize.

III. Analyze the Data: To analyze the data, students are asked to determine reasons for variability in the results of the random samples. They are also asked to make a dot plot of the 8 counts on the given line.

IV. Interpret the Data: Students are asked to find an approximate sampling distribution and use that information to find their best estimate of the total number of students wearing the school color on that given day.

Possible Extensions
1) As an extended project, students can collect data individually and share the results. Students will add the new findings to the original dot plot and see if how their new data affects their best estimate in part (e).
2) Small groups of students need to create a scenario by coming up with a question (could be school colors or they could explore a different question) where they can each survey the same number of students. Each will collect and share their data with the group. They will then collectively make the dot plot and examine the variation in the plot.
3) Use the activity as a formative assessment for sampling variability.
4) Have students create a random sample at your school. By obtaining a list of all students attending the school in an Excel spreadsheet. Make this spreadsheet available to students via computer in the classroom or school computer lab. Have each student in the class generate 100 random numbers using a graphing calculator, software, or random number table. The teacher will then construct spreadsheets with the names of the sampled student in each sample (note: some students may be sampled in more than one sample). Compile all of the posed questions in a survey to distribute to the sampled students at the school. Distribute the survey to all teachers at the school with a list of students that need to answer the questions. Results are collected the following day. This extension is given as Activity 2 below.

References
1. *Guidelines for Assessment and Instruction in Statistics Education* (GAISE) *Report*, ASA, Franklin et al., 2007 http://www.amstat.org/education/gaise/.

Activity Sheet 1
All of the 2300 students in your high school were asked to wear the school color today. The students who do will be seated in a special section of the bleachers for a last period rally. You want to estimate the total number of students who are wearing the school color today so you can mark off enough space in the bleachers.

To do this, you will take a random sample of 100 students during first period today and count the number of students that are wearing your school color.

a. In your random sample of size 100, you count 67 students who are wearing the school color. What is your best estimate of the total number of students in your high school who are wearing the school color?

b. You didn't know it, but seven other students independently had the same idea. They each took a random sample of students during first period and counted the number of students who are wearing the school color. Here are their results: 60, 64, 64, 66, 66, 68, 69.

Which of the following is **not** a cause for variability in these counts?

A. Some students could have miscounted.
B. It's difficult to tell what the school color is and what is a different shade of the color.
C. The eight students had different random samples.
D. Some students may have been in more than one sample.

c. Make a dot plot of these counts, including your own count of 67.

_____

d. The plot you made in c is an example of an approximate sampling distribution. Explain why the word "approximate" is used here.

e. Using the information in the approximate sampling distribution, what is the center of your plot? Explain how you are arriving at your answer.

Is your estimate from part (a) different than the center? How far off were you?

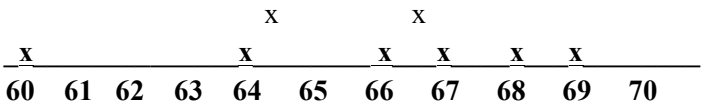
f. Using the plot you made in part c, would you say the values of the counts varied a lot or a little? Explain how you are arriving at your answer.

Answer Key

a. The best estimate of the total number of students in the high school that are wearing the school color is 67 out of every 100 or 67%.

b. Which of the choices is NOT a reason for variability in the given counts? D. This is because to construct an approximate sampling distribution, the repeated sampling is done with replacement. In other words, elements in a population could be selected in different samples.

c.



d. Approximate is used due to the fact that we only have 8 samples given, not ALL of the possible random samples.

e. The center of the distribution plot is about 66% because there are about the same amount of sample proportions on either side of it. However, because the lower point 60% is much lower than the highest point of 69%, in comparison to 66%, the lower point might actually bring down the center to about 65%.

f. The data varies about 9% points. Given the context of these data, the variation seems large. In other words, saving seats for a group of 60% would be much less than the number of seats saved for a group of 70%. In this sense, the variation seems to be large.

Activity Sheet 2

Part 1: Formulating Questions

Complete the given questions using the appropriate vocabulary:

1. Pose a question that you would be interested in finding out about your school population.

2. What is the population of interest for the question?

3. What is the parameter of interest for the question?

Part 2: Playing with Real Data

The following questions were asked at a high school of 600 students under the leadership of Principal Linz.

a. Do you participate in any school-sponsored extracurricular activities?
b. Do you have any siblings at the middle school or high school grade level?
c. Did you eat breakfast this morning?
d. Did you go to bed before 11:00pm last night?
e. Do you have a computer with Internet access at home?

A total of 493 students answered the survey.

For each of the survey questions, state the population of interest and the parameter of interest.

Suppose you took a sample from the 493 students and computed the statistic of interest.

For a given parameter, how would you be able to tell which statistics from our population are common and which are unusual?

How can we decide whether the following claims are plausible?

a. Extra-curricular activities are an integral part of a student's high school life. Furthermore, the participation in activities has been linked to student success (NCES, 1995). The federal Department of Education claims that 50% of high school students in the US participate in extra-

curricular activities. A high school principal surveys his students and find that 48% of students at his school participate in activities.

How could you decide whether the Department of Education's claim was plausible? Explain your answer.

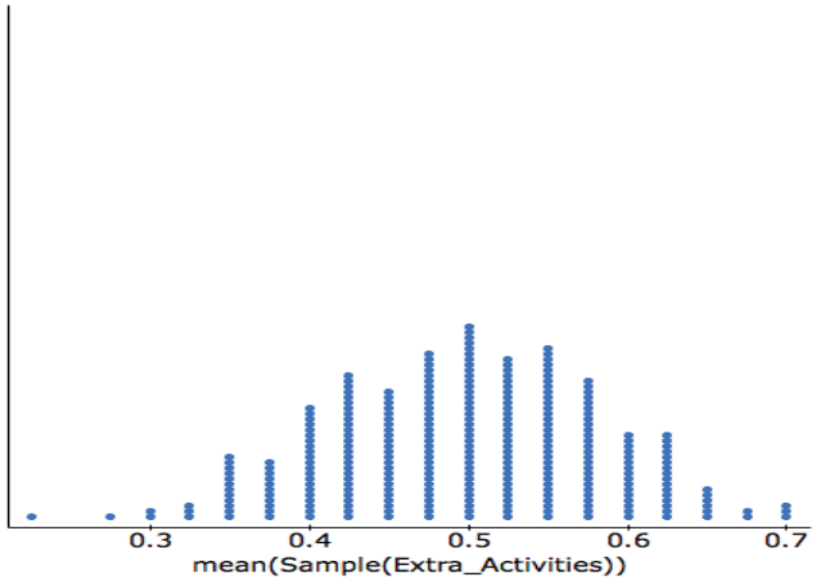
b. Eating and nutrition, in particular eating a good breakfast, has been linked to better student engagement and classroom learning. Unfortunately, the Department of Education claims that only 42% of students nationwide eat breakfast provided for them in the home.Principal Linz believes that her rate of breakfast eaters are much higher than 42%, as she is in a high-income area and has little documentation that her students have free and reduced lunch tickets.

   How would you decide whether the Department of Education's claim is plausible? Explain your answer.

Here is a sampling distribution for the proportion of kids in the high school that participate in extra-curricular activities constructed using samples of size 40.

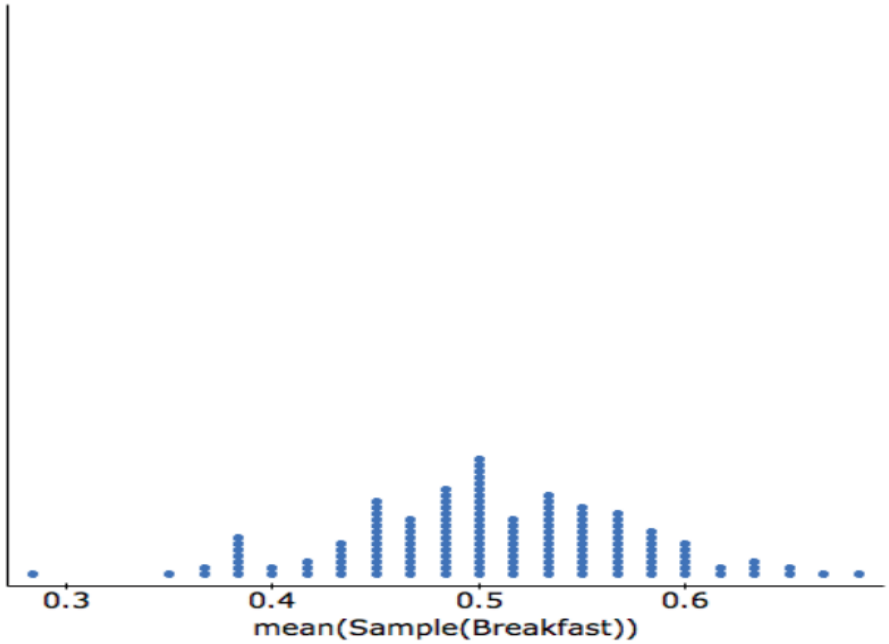
a. Using the School Data, take 1 random sample of size 40 yourself. Compute your summary statistic of interest.
b. Looking at the approximate sampling distribution, locate your summary statistic on the distribution. Mark it on the distribution.

c. Decide whether the summary statistic is common or unusual. How does this information relate to the claims above?



Here is a sampling distribution for the proportion of kids in the high school that eat breakfast constructed using samples of size 60.

d. Using the School Data, take 1 random sample of size 60. Compute your summary statistic of interest.
e. Looking at the approximate sampling distribution, locate your summary statistic on the distribution. Mark it on the distribution.
f. Decide whether the summary statistic is common or unusual. How does this information relate to the claim in 3 from part 1 above?


0.3
0.4
0.5
0.6
mean(Sample(Breakfast))