Using Dice to Introduce Sampling Distributions

Mary Richardson
Grand Valley State University
richamar@gvsu.edu

Published: July 2012

Overview of Lesson

In this activity students explore the properties of the distribution of a sample proportion under repeated random sampling. The population proportion of interest is the proportion of die rolls that land with a '6' on the up face. Students generate and analyze die roll data. The data analysis helps lead to an introduction of the sampling distribution of the sample proportion.

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.
7. Look for and make use of structure.
8. Look for and express regularity in repeated reasoning.

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. 2. Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, standard deviation) of two or more different data sets.
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.

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:
- 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:
- 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
Prior to completing this activity students have studied descriptive statistics and graphs, normal distributions, and basic probability concepts.

Learning Targets
Students will understand that a sample proportion is a sample statistic whose value will vary from one sample to another. Further, students will understand that the sampling distribution describes the probability distribution of a sample statistic.

Time Required
1 class period.

Materials Required
A copy of the Activity Worksheet; a graphing calculator; one or two dice per student.

Instructional Lesson Plan

The GAISE Statistical Problem-Solving Procedure

I. Formulate Question(s)
Circulate an Activity Worksheet (page 9) and one or two dice to each student. Ask students: if you were to roll a die an extremely large number of times, what proportion of the rolls will land with a '6' on the up face? Of course, they will answer 1/6. Explain to students that the goal of this activity is to use simulation to estimate the proportion of die rolls that would result in a '6' and then to examine the distribution of sample proportion values in repeated sampling.

II. Design and Implement a Plan to Collect the Data

Have students work individually using the data collection procedure described on the Activity Worksheet. Students determine their sample proportion of die rolls that land with a '6' on the up face for each of the sample sizes of $n = 10$, 20, and 60. Individual results are recorded on the Worksheet and each student writes her individual results on the whiteboard in an appropriately labeled column.

For a 30-student class there will be 30 sample proportion values for each of the sample sizes. Collecting this data provides an opportunity for students to participate in an example of obtaining repeated samples. Calculating the proportion of '6''s rolled for each of the sample sizes helps to reinforce the idea of a sample proportion being a random variable whose value changes from sample to sample.

After the individual sample proportions have been calculated and the results copied onto the white board, ask students to input the class data into the Class Data Table on the Activity Worksheet.

III. Analyze the Data

Figure 1 provides example class results. On this class day, 27 students participated in producing the example data set. Stem-and-leaf plots have been constructed for the proportions of dice landing on '6' for each of the 3 sample sizes. By examining the class data and answering a series of questions, students will begin to discover properties of the distribution of a sample proportion.

Stem-and-Leaf Plot for sample sizes of n=10

```
Frequency    Stem &  Leaf
   4.00       0 . 0000
    .00       0 .
   6.00       1 . 000000
    .00       1 .
  13.00       2 . 0000000000000
    .00       2 .
   4.00       3 . 0000
```

Stem width: .10
Each leaf: 1 case(s)

Stem-and-Leaf Plot for sample sizes of n=20

```
Frequency    Stem &  Leaf
    .00       0 .
   4.00       0 . 5555
   5.00       1 . 00000
   4.00       1 . 5555
   8.00       2 . 00000000
   1.00       2 . 5
   5.00       3 . 00000
```

Stem width: .10
Each leaf: 1 case(s)

Stem-and-Leaf Plot for sample sizes of n=60

Frequency   Stem & Leaf
   2.00   0 . 88
   3.00   1 . 002
   9.00   1 . 557778888
  10.00   2 . 0002333333
   3.00   2 . 557

Stem width: .10
Each leaf: 1 case(s)

Figure 1. Stem and leaf plots of example class data.
In the analysis that follows, the example class data in Figure 1 is used for purposes of illustration.

For each sample size students construct a stem-and-leaf plot of the sample proportion values and describe the shape, center, and spread of the distribution of values. For samples of size 10, there are only a few different values of the sample proportion, so the shape is difficult to determine. It appears that the center of the distribution is at around .20. And, the sample proportion values range from 0 to .30. For samples of size 20, more different values are obtained for the sample proportion of '6''s. The distribution almost appears to be somewhat uniform (rectangular). The center is once again at around .20. And, the sample proportions range from .05 to .30. For samples of size 60, the distribution of the sample proportions appears roughly mound-shaped (with a slight leftward skew). The center is at around .18. The sample proportion values range from .08 to .27.

For each sample size, students calculate the mean and standard deviation of the sample proportions. The calculated values of the mean of the sample proportion distributions are .16, .17, and .18, respectively (for samples of size 10, 20, and 60). The calculated values of the standard deviation of the sample proportion distributions are .09, .08, and .05, respectively (for samples of size 10, 20 and 60).

IV. Interpret the Results
Ask students to speculate about the relationship between the center of the distribution of the sample proportions and the value of the population proportion.   Students should note that the distribution of sample proportion values is centered on the value of the population proportion (1/6 is approximately .17).

Ask students to speculate about the relationship between the sample size and the shape of the distribution of the sample proportion. Students should note that as the sample size increases, the distribution of the sample proportion tends more towards a normal distribution.

Ask students: For which sample size is the standard deviation of the sample proportion values the largest and for which sample size is the standard deviation the smallest? Why do you suppose this happens? Students will note that the variability of the sample proportion values is related to the sample size, with a larger sample size resulting in smaller variability in the sample proportion values.

The results from the analyses of the repeated sampling can lead into a discussion about the theoretical properties of the sampling distribution of a sample proportion. The mean value of the distribution of a sample proportion for repeated random samples of size n drawn from the same population is equal to the corresponding value of the proportion of the population, p. The standard deviation of the distribution of a sample proportion for repeated random samples of size n drawn from the same population decreases with an increase in the sample size. The theoretical standard deviation formula $\sqrt{\dfrac{p(1-p)}{n}}$ can be introduced. Provided that the sample size is large enough (usually $np \geq 10$ and $n(1-p) \geq 10$), the distribution of a sample proportion for repeated random samples of size n drawn from the same population will be approximately normal (bell-shaped).

Contact **Author** for permission to use materials from this STEW lesson in a publication

Assessment

1. A polling organization polls 100 randomly selected registered voters to estimate the proportion of a large population that intends to vote for Candidate Y in an upcoming election. Although it is not known by the polling organization, the actual proportion of the population that prefers Candidate Y is .55.

(a) Give the numerical value of the mean of the sampling distribution of $\hat{p}$.

(b) Calculate the standard deviation of the sampling distribution of $\hat{p}$.

(c) If the proportion of the population that prefers Candidate Y is .55, would a sample proportion value of .70 be considered unusual?

2. Every student taking elementary statistics at a large university (about 1,100 students) participated in a class project by rolling a 6-sided die 100 times. Each individual student determined the *proportion* of his or her 100 rolls for which the result was a "1". The instructor plans to draw a histogram of the 1,100 sample proportions.

(a) What will be the approximate shape of this histogram?
A. Skewed left.
B. Uniform.
C. Normal (bell-shaped).
D. Skewed right.

(b) What will be the approximate mean for the 1,100 sample proportions?
A. 1/100
B. 1/6
C. 6/100
D. 6

(c) What will be the approximate standard deviation for the 1,100 sample proportions?

A. $\sqrt{\dfrac{(1/6)(5/6)}{1,100}}$

B. $\sqrt{\dfrac{(1/6)(5/6)}{100}}$

C. $\sqrt{\dfrac{(1/100)(99/100)}{1,100}}$

D. $\sqrt{\dfrac{(1/100)(99/100)}{100}}$

3. A candy factory makes 20% ($p = 0.20$) of all its candies with chocolate liquor. A random sample of $n = 100$ candies is taken, and $\hat{p}$ = proportion in the sample made with chocolate liquor. What is the z-score for $\hat{p} = .35$?

Contact Author for permission to use materials from this STEW lesson in a publication

Answers:
1.
(a) .55

(b) $\sqrt{\dfrac{.55(1-.55)}{100}} \approx .0497$

(c) Yes, a sample proportion of .70 would be considered unusual. It is roughly 3 standard deviations above the mean: $.55 + 3(.05) = .70$.

2.
(a) C
(b) B
(c) B

3. $z = \dfrac{\hat{p} - p}{\text{s.d.}(\hat{p})} = \dfrac{\hat{p} - p}{\sqrt{\dfrac{p(1-p)}{n}}}$ and, here: $p = .20$, $\hat{p} = .35$, and $n = 100$, so:

$z = \dfrac{.35 - .20}{\sqrt{\dfrac{.20(1-.20)}{100}}} = 3.75$

Possible Extension
After students have been introduced to the formula for a confidence interval for the population proportion, data collected from the dice activity for $n = 60$ can be used. Students can construct a confidence interval for the proportion of die rolls resulting in a '6'. Note that since each student's sample will be unique, after constructing their confidence interval, students can be asked to put their result onto the white board. This way, the instructor can have a discussion about what the confidence level means (examining the percentage of the different confidence intervals that include 1/6).

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. Adapted from: Gabrosek, J., Stephenson, P., and Richardson, M. (2007). How LO can you GO? Using the Dice-Based Golf Game GOLO to Illustrate the Sampling Distribution of a Sample Proportion. *Mathematics Teacher*, Volume 101, Issue 7.

3. Assessment questions taken from: Utts, J. and Heckard, B. (2007). *Mind on Statistics*, 3rd Edition. Thomson.

Using Dice to Introduce Sampling Distributions Activity Sheet

Suppose that you were to roll a regular six-sided die a large number of times. What proportion of the rolls will land with a '6' on the up-face? $p =$ _____.

Individual Data Table **Use two decimal places.**

	10 Trials	20 Trials	60 Trials
Number of Rolls Resulting in 6			
Proportion of Rolls Resulting in 6			

Copy your sample proportions (use two decimals) onto the white board in the appropriately labeled column. Input the class proportion of rolls resulting in a '6' into the Class Data Table.

Class Data Table Class *Proportion* of Rolls Resulting in a '6'.

Sample	$n = 10$	$n = 20$	$n = 60$	Sample	$n = 10$	$n = 20$	$n = 60$
1				16			
2				17			
3				18			
4				19			
5				20			
6				21			
7				22			
8				23			
9				24			
10				25			
11				26			
12				27			
13				28			
14				29			
15				30			

Use the Class Data to answer the following questions. Remember that $p = 1/6$.

1. For each sample size construct a stem-and-leaf plot of the sample proportion values and describe the shape, center, and spread of the distribution of values.

2. Based on your stem-and-leaf plots, what do you think is the relationship between the center of the distribution of the sample proportions and the value of the population proportion?

3. Based on your stem-and-leaf plots, what do you think is the relationship between the sample size and the shape of the distribution of the sample proportion?

4. For each sample size calculate the mean and standard deviation of the sample proportions and write a sentence to interpret the standard deviation.

Sample Size	Mean	Standard Deviation
$n = 10$		
$n = 20$		
$n = 60$		

5. For which sample size is the standard deviation the largest and for which sample size is the standard deviation the smallest? Why do you suppose this happens?