Lesson 16: Margin of Error when Estimating a Population Proportion

Student Outcomes

- Students use data from a random sample to estimate a population proportion.
- Students calculate and interpret margin of error in context.
- Students know the relationship between sample size and margin of error in the context of estimating a population proportion.

Lesson Notes

From prior lessons, students should recognize that the values of statistics calculated from samples selected from known populations will vary from sample to sample. In this lesson, students learn what variability can tell you about an unknown population. Students use data from a random sample drawn from a mystery bag to estimate a population proportion and then find and interpret a margin of error for the estimate. Comparing an observed proportion of successes from a random sample drawn from a population with an unknown proportion of successes to these sampling distributions provides information about what populations might produce a random sample like the one observed. This lesson can be done as a class investigation with students actually drawing samples and simulating sampling distributions using random-number generators or by working through a set of questions that relate to the kinds of activities they would actually do in a hands-on investigation.

Classwork

Exploratory Challenge 1 (25 minutes): Mystery Bag

For a Whole Class Hands on Investigation:

Prepare at least nine small bags of colored chips, one bag for each pair of students. If chips are not available, use pieces of paper with the letter "R" written on them. Each bag should have 20 chips with the following numbers of red chips. If the class is large, prepare duplicate bags for the percentages.

2 red chips – 10% of the chips are red
4 red chips – 20% of the chips are red
6 red chips – 30% of the chips are red
8 red chips – 40% of the chips are red
10 red chips – 50% of the chips are red
12 red chips – 60% of the chips are red
14 red chips – 70% of the chips are red
16 red chips – 80% of the chips are red
18 red chips – 90% of the chips are red

To learn about an unknown population, it is easiest to start by understanding how samples from a known population would behave. You might ask students the following question to activate their prior knowledge and remind them of the variability inherent in different samples from the same population.

MP.3

- Suppose you know that 20% of the chips in a bag are red. Write down an estimate of the number of red chips you are likely to see in a random sample of 30 chips from the bag. Have students write/speak responses to each other or as a class.

Drawing a red chip will constitute a "success." The proportion of red chips in each bag should be clearly written on the bottom of the bag. (Several pairs of students may have the same proportion, if the class is large.) The other chips can be any color other than red. These chips will represent "failures."

Give each pair of students a bag with the proportion of reds marked on the bottom. Monitor the class and when most students have two sets of 30 observations, bring the class together and suggest they use technology to generate their random samples to speed up the process. Each pair of students should generate a set that corresponds to the percentage of red chips in their bag; for example, the 10% can be represented by a set consisting of $s = \{1,0,0,0,0,0,0,0,0,0\}$, 20% by $\{1,1,0,0,0,0,0,0,0,0\}$, and so on. They should draw a sample of size 30 with replacement from their set. For some random-number generators, the command would be randsamp $(s, 30)$ to generate 30 elements from the set, and then a sum command can be used to find the number of successes in the sample:

s: = $\{1,0,0,0,0,0,0,0,0,0\}$ ▸ $\{1.,0.,0.,0.,0.,0.,0.,0.,0.,0.\}$

r: = randSamp(s,30) ▸ $\{0.,0.,0.,1.,0.,1.,0.,0.,0.,0.,0.,0.,1.,0.,0.,1.,0.,0.,0.,0.,0.,0.,0.,1.,0.,0.,0.,0.,0.,0.\}$

sum(r) ▸ $5.$

Have students record the number of successes in a frequency tally and then select additional samples. Repeating the process about 40 or 50 times will give them a simulated sampling distribution of the number of red chips in random samples of size 30 drawn from a population that is known to have a certain percentage of red chips. Emphasize that they are drawing samples from a population with KNOWN proportions of successes – in this case red chips.

Once most students have generated about 50 random samples of size 30 from their bags and recorded the number of reds in each sample in a frequency tally, bring the class together. Instruct students to do the following:

- Look at your simulated sampling distributions for the number of red chips in your samples, and write down an interval that seems to describe the number of reds they would typically get for the proportion of red chips in the bag.
  - *If I create a frequency tally or dot plot, it looks like about half of the samples resulted in having between* 10 *and* 14 *red chips.*
- Then return to the mystery bag and have students respond to Exercises 2 and 3.

For a Lesson With and Without the Hands-On Activities:

Prepare one "mystery bag" that has 20 chips with an unknown (to the class) proportion of red chips. You might put eight red chips in the bag for a mystery proportion of 40%. Try not to use six red chips (30%) as students may confuse the percent with the sample of 30 chips that will be drawn with replacement from the bag. Students do not actually have to know there are 20 chips in any of the bags because the samples will be drawn with replacement.

MP.2 Begin the class with the introduction below. Lessons without the hands-on activities should begin with Exercise 1. Lessons with the hands-on activities should begin with Exercise 2. In Exercise 3, students understand that a margin of error is the interval that marks off the proportions of red chips from the expected proportion that are unlikely to occur based on the simulated sampling distribution. If the proportion from the sample did not show up in the sampling distribution or was one of the more extreme proportions identified in the interval called the margin of error, then it is unlikely the proportion of red chips in the mystery bag is equal to the proportion stated in the investigation. It will be important to have a class discussion on part (b) of Exercise 3 to make sure students understand that the margin of error defines an interval and not the likelihood of making a mistake.

Bring in the mystery bag and ask students what proportion of the chips in the bag they think are red chips. Ask students:

MP.3
- How can you find the proportion of red chips in the bag? Students should write or speak with a partner to develop a plan. (Note: Taking the chips out, examining them, and counting the red ones is not an option.)

Have one student draw a chip from the bag and a different student record whether the chip was red or not red. Return the chip to the bag, shake the bag, and have the students draw and record the color of a second chip. Continue the process until they have a sample of 30 chips. Ask students to write down their predictions for the proportion of red chips in the bag based on the sample results.

Scaffolding:

Encourage advanced learners to develop their own plans for determining the proportion and carry it out, without scaffold questions given.

Offer struggling students a simpler example (e.g., a bag with only four chips—one red—in it) that illustrates the ideas at work here. Show a visual of this and ask questions such as, "How many red chips would you have in a sample of 10? 20? 50?"

Exercises 1–4

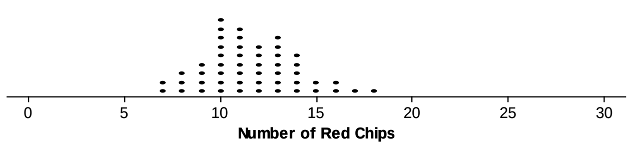
Exercises 1–4

In this lesson you will use data from a random sample drawn from a mystery bag to estimate a population proportion and learn how to find and interpret a margin of error for your estimate.

1. **Write down your estimate for the proportion of red chips in the mystery bag based on the random sample of 30 chips drawn in class.**

   Possible answer: If 15 red chips were in the sample of 30, some students might suggest that $\frac{15}{30}$ (or 0.5) of the chips in the mystery bag were red. Others might suggest an interval around 0.5.

2. **Tanya and Raoul had a paper bag that contained red and black chips. The bag was marked 40% red chips. They drew random samples of 30 chips, with replacement, from the bag. (They were careful to shake the bag after they replaced a chip.) They had nine red chips in their sample. They drew another random sample of 30 chips from the bag, and this time they had 12 red chips. They repeated this sampling process 50 times and made a plot of the number of red chips in each sample. A plot of their sampling distribution is shown below.**



   a. **What was the most common number of red chips in the 50 samples? Does this seem reasonable? Why or why not?**

      Possible answer: The most common number in the samples was 10 red chips, which seems reasonable because we would expect to have about 40% successes, and 40% of 30 is 12 successes and 10 is close to that.

b. **What number of red chips, if any, never occurred in any of the samples?**

They never got a sample with a number of red chips less than or equal to 6 or more than or equal to 20.

c. **Give an interval that contains the "likely" number of red chips in samples of size 30 based on the simulated sampling distribution.**

Possible answer: From $7-19$ red chips.

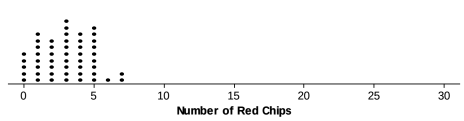
Note: do not focus on exactly what "likely" means. The object is to see if it is at all reasonable for an outcome to occur by chance.

d. **Do you think the number of red chips in the mystery bag could have come from a sample drawn from a bag that had 40% red chips? Why or why not?**

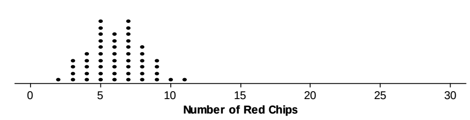
Possible answer: This depends on the number of red chips that were observed in the random sample drawn from the mystery bag. If the observed number of red chips was 18, the answer would be "yes" because 18 red chips occurred just by chance in the simulated sampling distribution. Some students might suggest 18 or more red chips only occurred twice in 50 random samples so it was not too likely but could happen by chance. The important concept is that students look at the simulated distribution to see where the observed outcome falls with respect to that distribution.

Nine different bags of chips were distributed to small teams of students in the class. Each bag had a different proportion of red chips. Each team simulated drawing 50 different random samples of size 30 from their bag and recorded the number of red chips for each sample. The graphs of their simulated sampling distributions are shown below.

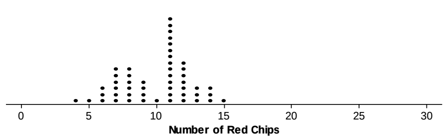
Population 10% red chips



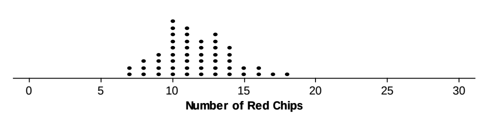
Population 20% red chips



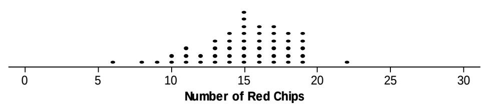
Population 30% red chips



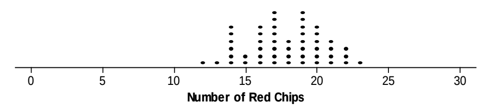
Population 40% red chips



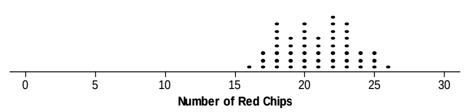
Population 50% red chips



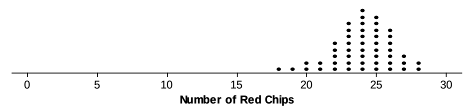
Population 60% red chips



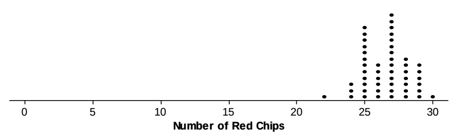
Population 70% red chips



Population 80% red chips



Population 90% red chips



3. Think about the number of red chips in the random sample of size 30 that was drawn from the mystery bag.

   a. Based on the simulated sampling distributions, do you think that the mystery bag might have had 10% red chips? Explain your reasoning.

   Possible answer: If the random sample of size 30 from the mystery bag had 18 red chips, the answer would be "no" because 18 never showed up once in all of the samples.

   b. Based on the simulated sampling distributions, which of the percentages 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, and 90% might reasonably be the percentage of red chips in the mystery bag?

   Possible answer: If the number of red chips in the sample from the mystery bag was 18, it looks like, just by chance, the sample could have been drawn from a population having from 40% to 80% red chips.

   c. Let p represent the proportion of red chips in the mystery bag. (For example, $p = 0.40$ if there are 40% red chips in the bag.) Based on your answer to part (b), write an inequality that describes plausible values for p. Interpret the inequality in terms of the mystery bag population.

   Possible answer: $0.40 \leq p \leq 0.80$. This means that based on the simulated sampling distributions, the true proportion of red chips in the mystery bag could have been anywhere from 0.40 to 0.80. It would not have been surprising for a random sample of size 30 drawn from any of these populations to have included 18 red chips.

4. If the inequality like the one you described in part (c) of Exercise 3 went from 0.30 to 0.60, it is sometimes written as 0.45 ± 0.15. The value 0.15 is called a "margin of error." The margin of error represents an interval from the expected proportion that would not contain any proportions or very few proportions based on the simulated sampling distribution. Proportions in this interval are not expected to occur when taking a sample from the mystery bag.

a. **Write the inequality you found in Exercise 3, part (c) using this notation. What is the margin of error?**

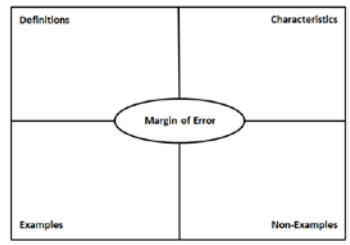
Possible answer: Using 18 as the number of red chips in the random sample from the mystery bag, the interval would be 0.60 ± 0.20. The margin of error is defined as an interval of 0.20.

b. **Suppose Sol said, "So this means that the actual proportion of red chips in the mystery bag was 60%." Tonya argued that the actual proportion of red chips in the mystery bag was 20%. What would you say?**

Possible answer: They are both wrong. The notation does not mean that the center of the interval is the actual population but that it is the center of an interval made by adding and subtracting the margin of error. A random sample drawn from any proportion in the interval could have produced an outcome of 18 red chips. Tonya has mixed up the number describing the length of the interval with the population proportion.

Scaffolding:

- ELL learners may need a quick explanation of the term "margin."
- It can be the part of a page that is above, below, or to the side of a printed part.
- In this lesson, it can be a measure or degree of difference.
- A Frayer diagram may be used to explain margin of error.



Exploratory Challenge 2 (10 minutes): Samples of Size 50/Exercises 5–7

All students should do Exercise 5. As in the prior part of the lesson, you may want students (or a subset of students) to actually simulate the sampling distributions. Or, you can have them use the distributions provided in Exercise 6. The simulation would be similar to that for a sample of size 30, simply replacing 30 in the command with 50. All students should respond to Exercise 6.

Exploratory Challenge 2: Samples of Size 50/Exercises 5–7

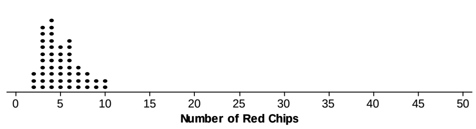
MP.3

5. **Do you think the "margin of error" would be different in Exercise 4 if you had sampled 50 chips instead of 30? Try to convince a partner that your conjecture is correct.**

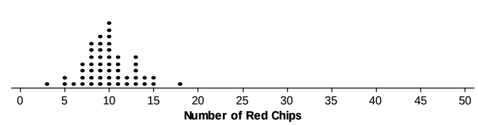
Possible answer: They might be different but maybe only a little bit. I'm not sure why the sample size would make a difference because the counts would be different, but it would still be centered around the same proportion (40% of a sample of size 30 is 12 red chips; 40% of a sample of size 50 is 20 red chips – different counts but the same proportion).

6. **Below are simulated sampling distributions of the number of red chips for samples of size 50 from populations with various percentages of red chips.**

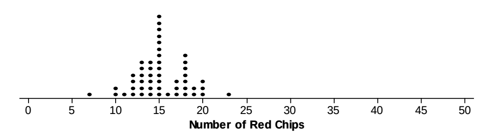
Population with 10% red chips



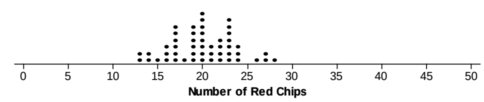
Population with 20% red chips



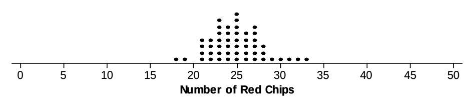
Population with 30% red chips



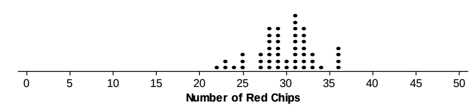
Population with 40% red chips



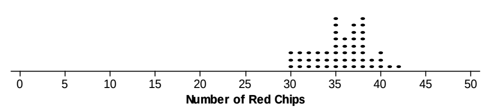
Population with 50% red chips



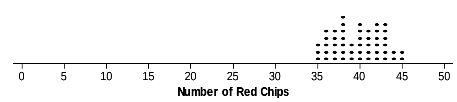
Population with 60% red chips



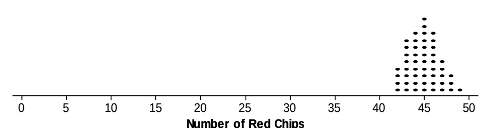
Population with 70% red chips



Population with 80% red chips



Population with 90% red chips



a. **Suppose you drew 30 red chips in a random sample of 50 from the mystery bag. What are plausible values for the proportion of red chips in the mystery bag? Explain your reasoning.**

   Possible answer: Plausible population proportions are from 0.50 to 0.70.

b. **Write an expression that contains the margin of error based on your answer to part (a).**

   Possible answer: The margin of error statement would be 0.60 ± 0.10. The margin of error would be 0.10.

7. **Remember your conjecture from Exercise 5, and compare the margin of error you found for a sample of size 30 (from Exercise 3) to the margin of error you found for a sample of size 50.**

   a. **Was your reasoning in Exercise 5 correct? Why or why not?**

      Possible answer: I forgot that the sample size might have to do with the spread of the distribution not just the center, so my reasoning was not correct.

b. **Explain why the change in the margin of error makes sense.**

Possible answer: The margin of error was 0.10 for the sample size of 50, which is less than the margin of error for the sample size of 30. It makes sense that the margin of error would decrease as the sample size increases because as the sample size increases, the variability from sample to sample decreases, and the sample proportions tend to be closer to the actual population proportion.

Closing (5 minutes)

- Why do you suppose we use language like "margin of error" to define the interval describing a population that might have produced a sample proportion of red chips like the one from the mystery bag?
  - *Possible answer: The word "error" is misleading in that it makes you think that you are "off" by a certain percent; it really tells you the range of possible population proportions for the population from which a random sample is drawn.*
- How could you apply the concepts from this lesson to investigate the proportion of females in a coffee shop on a weekday morning if you had observed 18 females in a random sample of 30 people during that time?
  - *Possible answer: The results would be the same because females would be like the red chips, and the sample size remains the same. Going into the coffee shop to take a random sample is the same as drawing a random sample from a mystery bag.*
- You knew the sample size and the observed outcome when you started the investigation you just did. Why were these important to know?
  - *Possible answer: You needed the sample size to know what size samples to generate, and you had to have an observed outcome to think about what might be plausible populations considering the simulated distributions of sample proportions for populations of known proportions.*

Ask students to summarize the main ideas of the lesson in writing or with a neighbor. Use this as an opportunity to informally assess comprehension of the lesson. The Lesson Summary below offers some important ideas that should be included.

Lesson Summary

In this lesson, you investigated how to make an inference about an unknown population proportion based on a random sample from that population.

- **You learned how random samples from populations with known proportions of successes behave by simulating sampling distributions for samples drawn from those populations.**
- **Comparing an observed proportion of successes from a random sample drawn from a population with an unknown proportion of successes to these sampling distributions gives you some information about what populations might produce a random sample like the one you observed.**
- **These plausible population proportions can be described as $p \pm M$. The value of M is**

Exit Ticket (5 minutes)

Name ______________________________ Date ______________

Lesson 16: Margin of Error when Estimating a Population Proportion

Exit Ticket

1. Suppose you drew a sample of 12 red chips in a sample of 30 from a mystery bag. Describe how you would find plausible population proportions using the simulated sampling distributions we generated from populations with known proportions of red chips.

2. What would happen to the interval containing plausible population proportions if you changed the sample size to 60?

Exit Ticket Sample Solutions

1. **Suppose you drew a sample of 12 red chips in a sample of 30 from a mystery bag. Describe how you would find plausible population proportions using the simulated sampling distributions we generated from populations with known proportions of red chips.**

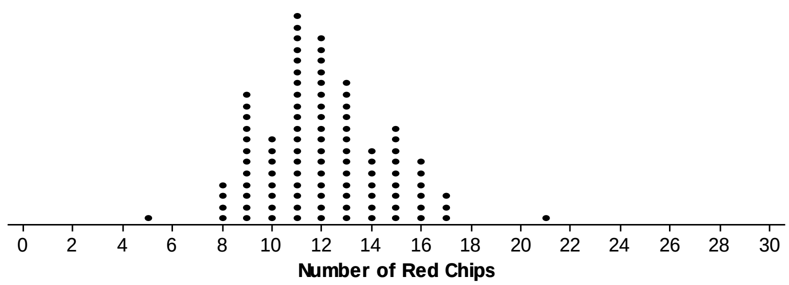
   Answers will vary. I would look at the simulated distributions to see which ones contained an outcome of 12 red chips. For those that did, the corresponding population proportion would be included in the set of plausible population proportions. (Students might use the actual distributions if they have them available, and use the corresponding interval as an example in their answers – 0.20 to 0.50 in those above.)

2. **What would happen to the interval containing plausible population proportions if you changed the sample size to 60?**

   Possible answer: The width of the interval would decrease, and the margin of error would be smaller.

Problem Set Sample Solutions

1. **Tanya simulated drawing a sample of size 30 from a population of chips and got the following simulated sampling distribution for the number of red chips:**



   Which of the following results seem like they might have come from this population? Explain your reasoning.

   I. 8 red chips in a random sample of size 30

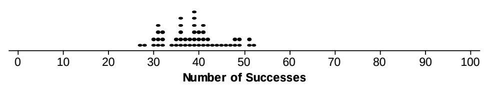
   II. 12 red chips in a random sample of size 30

   III. 24 red chips in a random sample of size 30

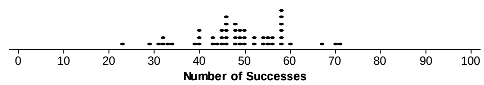
   Possible answer: Samples that had 8 and 12 red chips might have come from this population because they occurred by chance in the random samples from the simulation. A sample with 24 red chips never occurred by chance, so it seems more unlikely to happen for this population.

2. 64% of the students in a random sample of 100 high school students intended to go onto college. The graphs below show the result of simulating random samples of size 100 from several different populations where the success percentage was known and recording the percentage of successes in the sample.

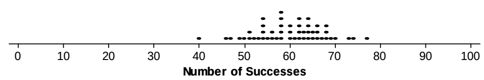
(i) Population with 40% successes



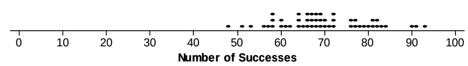
(ii) Population with 50% successes



(iii) Population with 60% successes



(iv) Population with 70% successes



a. Based on these graphs, which of the following are plausible values for the percentage of successes in the population from which the sample was selected: 40%, 50%, 60%, or 70%? Explain your thinking.

Possible Answer: 64% successes was a likely outcome for samples from populations with 60% and 70% successes. While exactly 64% did not occur in the 50% success population, it was in the range of observed sample percentages and, thus, could have happened. None of the samples from the 40% success population had a percentage of successes as large as 64%, so it would not seem likely that the sample came from this population.

b. Would you need more information to determine plausible values for the actual proportion of the population of high school students who intend to go to some postsecondary school? Why or why not?

Possible answer: Yes, you would ned more information because you have not really looked at any simulated distributions of sample proportions larger than 70%. And 80% or 90% might turn out to be plausible as well.

3. Suppose the mystery bag had resulted in the following number of red chips. Using the simulated sampling distributions found earlier in this lesson, find a margin of error in each case.

a. The number of red chips in a random sample of size 30 was 10.

Possible answer: 0.20 to 0.50 or 0.35 ± 0.15, for a margin of error of 0.15.

b. The number of red chips in a random sample of size 30 was 21.

Possible answer: 0.50 to 0.80 or 0.65 ± 0.15, for a margin of error of 0.15.

c. The number of red chips in a random sample of size 50 was 22.

Possible answer: 0.40 to 0.60 or 0.50 ± 0.10, for a margin of error of 0.10.

4. The following intervals were plausible population proportions for a given sample. Find the margin of error in each case.

   a. from 0.35 to 0.65

      $\mathbf{0.50 \pm 0.15}$

   b. from 0.72 to 0.78

      $\mathbf{0.75 \pm 0.03}$

   c. from 0.84 to 0.95

      $\mathbf{0.895 \pm 0.055}$

   d. from 0.47 to 0.57

      $\mathbf{0.52 \pm 0.05}$

5. Decide if each of the following statements is true or false. Explain your reasoning in each case.

   a. The smaller the sample size, the smaller the margin of error.

      False. The smaller the sample size, the larger the margin of error.

   b. If the margin of error is 0.05 and the observed proportion of red chips is 0.35, then the true population proportion is likely to be between 0.40 and 0.50.

      True. 0.40 to 0.50 is the range of plausible values for the population proportion.

6. Extension: The margin of error for a sample of size 30 is 0.20; for a sample of 50, is 0.10. If you increase the sample size to 70, do you think the margin of error for the percent of successes will be 0.05? Why or why not?

   No. When we simulated the sampling distributions for a sample size 100, the margin of error got smaller but was not 0.05.