Are You Ready for

Statistics

in

Algebra 2?

Robyn Poulsen

Lake Placid Middle/High School

T^3 Regional Instructor

NYS Master Teacher Fellow

AMTNYS Executive Board

(HS Level Rep)

HMVAMC

March 19, 2016

Niskayuna, NY

2 KEY IDEAS in STATISTICS
&
ALIGNMENT to Common Core

Key Idea	NY CC Standard
Confidence Interval Based on the data in a sample, find a range of plausible values for a quantity in a population.	**S.IC.B.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.
Hypothesis Test Is an effect observed in a sample true for a population or just due to random chance?	**S.IC.B.5:** Use data from a randomized experiment to compare two treatments; use simulations to decide if differences between parameters are significant.

THANK YOU FOR ATTENDING!

Please be in touch with any questions, comments, concerns, ideas. We can do this TOGETHER ☺

rpoulsen@lpcsd.org

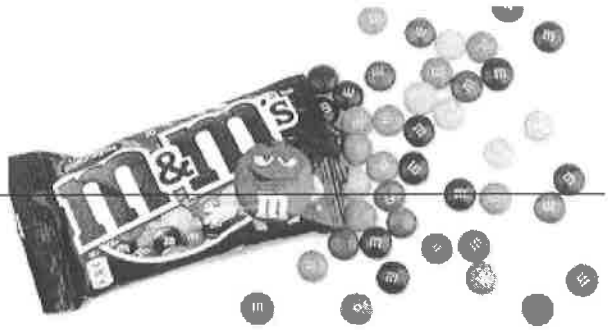
http://lpcsdrpoulsen.weebly.com/for-math-teachers.html

@RobynPoulsen

RESOURCES USED

- StatMonkey – Jason Molesky – Everything I ever needed to Know about AP Statistics I Learned from a Bag of m&m's
- Math = Love blog
- Stats Modeling the World 1st Edition - Bock, Velleman, DeVeaux
- The Practice of Statistics – Yates, Moore, Starnes
- New York Coach – Mastering The High School Standards Algebra II – Triumph Learning
- Big Ideas Math Algebra 2 – Ron Larson & Laurie Boswell
- engageNY.org – Algebra II Module 4 Parts B-D
- PARCC

A2CC @LP



PROBLEM SOLVING MONDAY

Everything I need to know about Statistics, I learned from a bag of m&m's ...

Welcome to statistics! Your task today is to collect data about bags of m&m's through a carefully designed and controlled experiment.

The steps today:

1. Get a bag of m&m's from Mrs. Poulsen (You're welcome! ☺)
2. Weigh the bag. Record the weight below.
3. Open the bag.
4. Count the number of candies in your bag.
5. Sort the m&m's by color.
6. Count the number of m&m's of each color and record your data accurately below.
7. Record your color distribution on the class dotplots.
8. Carefully and properly dispose of the m&m's by seeing whether or not they really do melt in your mouth, not in your hand.

SAVE THIS PAPER ~ WE WILL BE USING THIS DATA THROUGHOUT THE UNIT

DATA COLLECTED:

Weight of my bag: _____ Number of candies in my bag: _____

Color Distribution of candies in my bag:

Blue	Brown	Green	Orange	Red	Yellow

According to Mars Corporation (the makers of m&m's), the color distribution of plain m&m's candies is:

Blue	Brown	Green	Orange	Red	Yellow
24%	13%	16%	20%	13%	14%

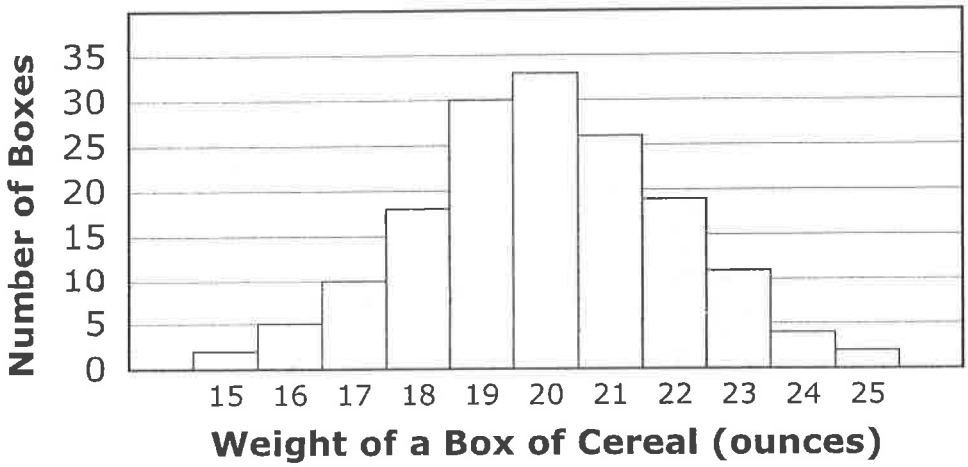
Convert the color distribution count of YOUR bag of candies to percentages. Compare them to the Mars Corporation claim above. Is there a discrepancy? Explain why you think this happens.

Blue	Brown	Green	Orange	Red	Yellow

Use the information provided to answer Part A and Part B for question 24.

The distribution of weights (rounded to the nearest whole number) of all boxes of a certain cereal is approximately normal with mean 20 ounces and standard deviation 2 ounces.

A sample of the cereal boxes was selected, and the weights of the selected boxes are summarized in the histogram shown.



24. Part A

If w is the weight of a box of cereal, which range of weights includes all of the weights of cereal boxes that are within 1.5 standard deviations of the mean?

Ⓐ $17 \leq w \leq 23$

Ⓑ $18.5 \leq w \leq 21.5$

Ⓒ $19 \leq w \leq 21$

Ⓓ $20 \leq w \leq 23$

Part B

Which of these values is the **best** estimate of the number of boxes in the sample with weights that are more than 1.5 standard deviations above the mean?

Ⓐ 2

Ⓑ 6

Ⓒ 17

Ⓓ 36

5. A researcher gathers data on how long teenagers spend on individual cell phone calls (in number of minutes). Suppose the research determines that these calls have a mean 10 minutes and standard deviation 7 minutes.

   a. Suppose the researcher also claims that the distribution of the call lengths follows a normal distribution. Sketch a graph displaying this distribution. Be sure to add a scale and to label your horizontal axis.

b. Using your graph, shade the area that represents the probability that a randomly selected call lasts more than 12 minutes. Is this probability closer to 0.50 or to 0.05?

c. After looking at the above mean and standard deviation of the call length data, a second researcher indicates that she does not think that a normal distribution is an appropriate model for the call length distribution. Which researcher (the first or the second) do you think is correct? Justify your choice.

17. The heights of the male students at a college are approximately normally distributed. Within this curve, 95% of the heights, centered about the mean, are between 62 inches and 78 inches. The standard deviation is 4 inches. Use this information to estimate the mean height of the males. Approximate the probability that a male student is taller than 74 inches. Explain how you determined your answers.

Enter your answers and your explanation in the space provided.

Use the information provided to answer Part A and Part B for question 33.

The manager of food services at a local high school is interested in assessing student opinion about a new lunch menu in the school cafeteria. The manager is planning to conduct a sample survey of the student population.

33. Part A

Which of the listed methods of sample selection would be the **most** effective at reducing bias?

Ⓐ Randomly select one day of the week and then select the first 30 students who enter the cafeteria on that day.

Ⓑ Post the survey on the school Web site and use the first 30 surveys that are submitted.

Ⓒ Randomly select 30 students from a list of all the students in the school.

Ⓓ Randomly select one classroom in the school and then select the first 30 students who enter that classroom.

Part B

The manager wants to know if a student's gender is related to the student's opinion about the menu. Which statement **best** describes the study?

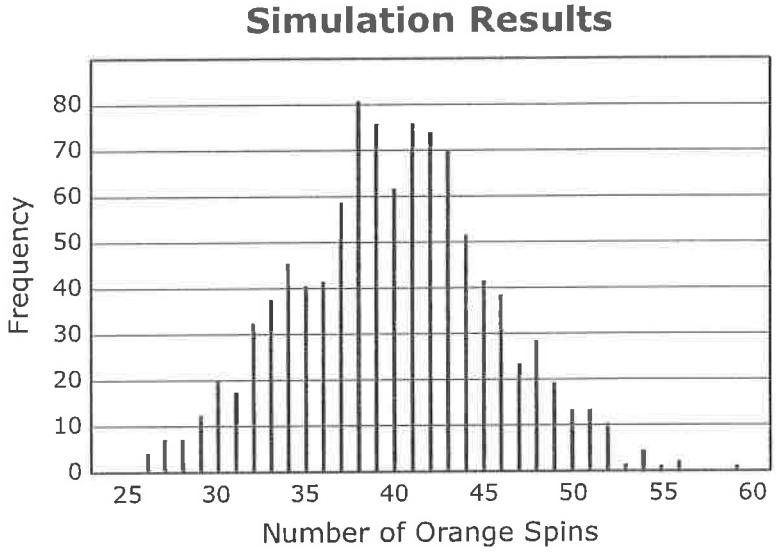
Ⓐ This is an observational study, and therefore, the manager will be able to establish a cause-and-effect relationship between gender and opinion.

Ⓑ This is an observational study, and therefore, the manager will not be able to establish a cause-and-effect relationship between gender and opinion.

Ⓒ This is an experimental study, and therefore, the manager will be able to establish a cause-and-effect relationship between gender and opinion.

Ⓓ This is an experimental study, and therefore, the manager will not be able to establish a cause-and-effect relationship between gender and opinion.

15. A circular spinner is divided into five sectors of different colors. A student spun the arrow on the spinner 200 times and recorded that the arrow stopped on the orange sector 38 times out of the 200 spins. To test whether the spinner was fair, the student used a computer to simulate the number of times the arrow stops on orange in 200 spins of a fair spinner equally divided into five sectors of different colors. The results of 1,000 trials of the simulation are shown.



Based on the results of the simulation, is there statistical evidence that the spinner is not fair?

Ⓐ Yes, because 38 was the most frequent outcome.

Ⓑ Yes, because about 8% of the outcomes were 38.

Ⓒ No, because the distribution is approximately normal.

Ⓓ No, because an outcome of 38 or less is not unusual.

Name _____ Date _____

1. Suppose you wanted to determine whether students who close their eyes are better able to estimate when 30 seconds have passed than students who do not close their eyes. (You ask students to tell you when to stop a stopwatch after they think 30 seconds have passed.) You find the first 50 students arriving at school one day. For those 50, you flip a coin to decide whether or not they will close their eyes during the test. Then, you compare the amounts by which each group overestimated or underestimated.

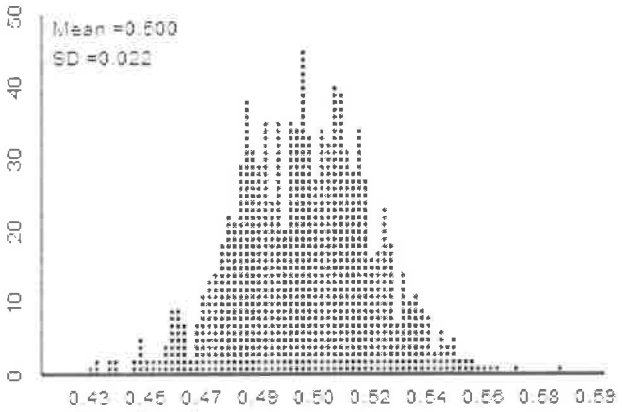
   a. Did this study use *random sampling*? Explain your answer by describing what purpose random sampling serves in such a study.

   b. Did this study make use of *random assignment*? Explain your answer by describing what purpose random assignment serves in such a study.

   c. Would the study described above be an *observational study* or an *experimental study*? Explain how you are deciding.

2. A Gallup poll conducted July 10–14, 2013, asked a random sample of U.S. adults: "How much attention do you pay to the nutritional information that is printed on restaurant menus or posted in restaurants, including calories and sugar and fat content?" The sample results were that 43% of the respondents said they pay a "fair amount" or a "great deal" of attention. Suppose there had been 500 people in the study.

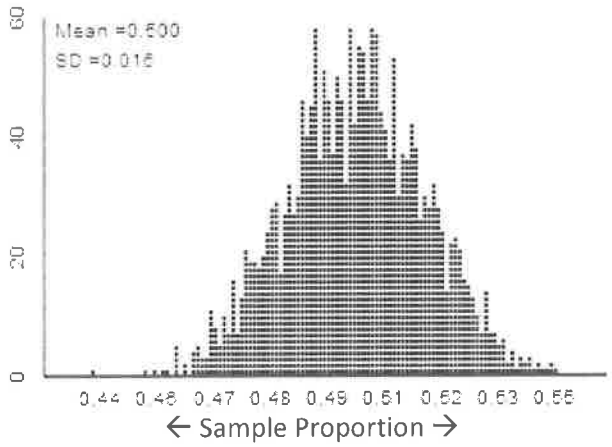
   The following graph displays the results from 1,000 random samples (each with sample size 500) from a very large population where 50% of respondents "pay some attention" and 50% "pay little or no attention."



   a. Based on the simulation results above, are the sample data (43% responding "pay some attention") consistent with the simulation? In other words, do these results cause you to question whether the population is 50/50 on this issue? Explain.

   Do you believe it is reasonable to generalize the results from this study to all U.S. adults? Explain.

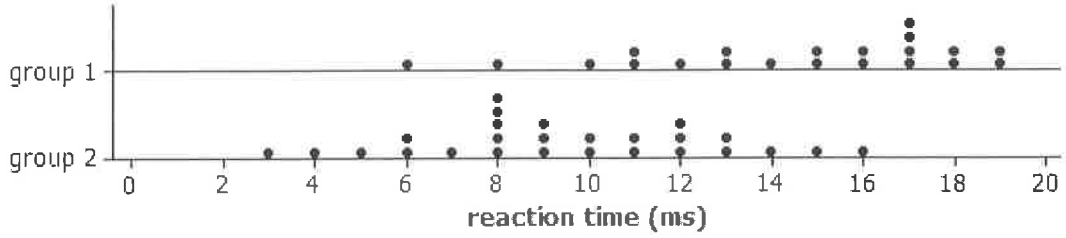
Suppose Gallup plans to conduct a new poll of a random sample of 1,000 U.S. adults on an issue where the population is evenly split between two responses. The following graph displays the results from 2,000 random samples (each with sample size 1,000) from such a population.



b. Based on these simulation results, estimate the expected margin of error for the Gallup poll. Explain how you developed your estimate.

c. Suppose the study used a sample size of 2,000 instead of 1,000. Would you expect the margin of error to be larger or smaller?

3. A randomized experiment compared the reaction time (in milliseconds) for subjects who had been sleep deprived (group 1) and subjects who had not (group 2).



Variable	N	Mean	StDev
group 1	21	14.38	3.64
group 2	26	9.50	3.33

a. Based on the above output, for which group would it be more reasonable to use a normal curve to model the reaction time distribution? Justify your choice.

b. The difference in means is $14.38 - 9.50 = 4.88$. One of the researchers claims that the reaction time if you are sleep deprived is 5 ms greater than the reaction time if you are not sleep deprived. Explain one reason why this claim is potentially misleading.

c. Describe how to carry out a simulation analysis to determine whether the mean reaction time for group 1 is significantly larger than the mean reaction time for group 2.

d. The graph below displays the results of 100 repetitions of a simulation to investigate the difference in sample means when there is no real difference in the treatment means. Use this graph to determine whether the observed mean reaction time for group 1 is significantly larger than the observed mean reaction time for group 2. Explain your reasoning.