APPENDIX A

Informed Consent

You are invited to participate in a doctoral research project entitled "Elementary Preservice Teachers' Conceptions of Variation", being conducted by Daniel Canada from the Department of Mathematical Sciences at Portland State University. The researcher hopes to develop a characterization of the knowledge held by elementary preservice teachers about this important statistical concept. You were selected as a possible participant by virtue of your enrollment in the Math 212 class.

By giving your consent to take part in this study, you are agreeing to three distinct aspects of data gathering. First of all, comments made by you in class which the researcher deems pertinent can be transcribed and used as data. Secondly, homework which is relevant to the project can be photocopied and used as data. Thirdly, you agree to participate in at least one interview which takes place outside of the normal class hours. The interview will be scheduled at a mutually convenient time and place; it will be videotaped, and will last approximately one hour. The transcripts from this interview can also be used as data.

You as a prospective teacher will gain a direct benefit from a deeper exploration of your own ideas about this key statistical concept; this exploration allows you to extend your own learning about variation in the non-evaluative environment of the research project. Moreover, the practice in articulating your thinking is especially helpful as you make the transition to your own classroom, and invoke similar practices with your own students.

Potential risks include the possibility that an unauthorized person may view the data, or that your actual name may inadvertently become associated with the data. To minimize this risk, all written responses, notes, audio and video tapes, and transcriptions will be kept confidential, and will be kept locked up in the researcher's office in the Department of Mathematical Science at PSU. After three years, these records will be destroyed. In writing any results for the study, pseudonyms will be used so that your identity cannot be matched with the responses you have provided. There is also a risk that having a researcher in the classroom may affect the learning environment. To lessen this risk, it will be stressed that this research is descriptive and not evaluative in nature.

Your participation in this study is voluntary and you are completely free to withdraw from the study at any time. Your decision to participate or not will not affect your relationship with the researcher or with any academic program at PSU in any way. If you have concerns about your participation in this study or your rights as a research subject, please contact the Human Subjects Research Review Committee, Office of Research and Sponsored Projects, 111 Cramer Hall, Portland State University, (503) 725-8182. If you have any questions about the study itself, please contact Daniel Canada, at the Department of Mathematical Sciences, 334 Neuberger Hall, Portland State University, (503) 725-3621.

Your signature indicates that you have read and understand the above information and agree to take part in this study. Please remember that you may withdraw your consent at any time without penalty. Also, by signing, you are not waiving any legal claims, rights or remedies. The researcher has provided you with a copy of this form for your records.

____
Signature of Participant Date

____
Daniel Canada, Researcher Date
Department of Mathematical Sciences
Portland State University
(503) 725-3621

APPENDIX B

<u>Surveys and Interviews</u>

The surveys and interview scripts are appended in the order that they were administered:

- PreSurvey
- PreInterview
- PostSurvey (Data & Graphs)
- PostSurvey (Sampling)
- PostSurvey (Probability)
- PostInterview

PRESURVEY

Mth 212 Survey on Probability and Statistics Name: ____________

1] Where & when did you take Math 211?

2] If you have taken prior math courses in which probability and/or statistics was taught,

a) When & where did you take those courses ?

b) How did you feel about the probability and/or statistics at that time?

3] How comfortable do you feel about learning probability and/or statistics now ?

4] What does the word “random” mean to you ?

Give an example of something that happens in “a “random” way.

5] What does the word “variation” mean to you ?

Give an example of something that “varies”.

This set of questions helps to give a picture of how you think about some problems in probability and statistics. Rather than think in terms of a right or wrong answer, just write down your best thinking for each situation. Later in the quarter, we'll explore situations like these as a class.

[1] Suppose there is a container with 100 pieces of candy in it. 60 are Red, and 40 are Yellow. The candies are all mixed up in the container.

You reach in and pull out a handful of 10 candies at random.

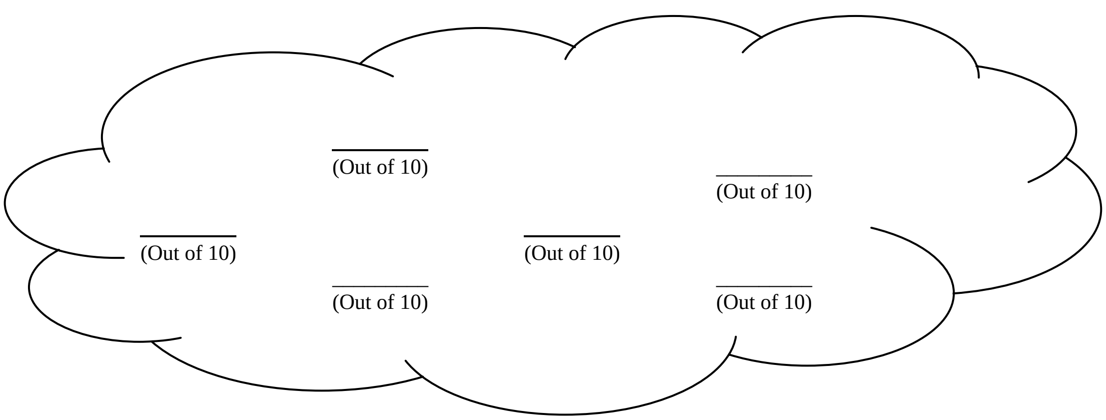
(a) How many red candies do you think you might get?

Why do you think this?

(b) Suppose you do this several times (each time returning the previous handful of 10 candies and remixing the container). Do you think this many reds would come out every time?

Why do you think this?

(c) Suppose six classmates do this experiment (each time returning the previous handful of 10 candies and remixing the container). Write down the number of reds that you think each classmate might get:



Why did you choose those numbers?

[2] Suppose 6 people did this experiment – pulled ten candies from the container, wrote down the number of reds, then returned the ten and remixed all the candies.

What do you think the numbers of reds will most likely go from?

From a low of ______ to a high of _______.

Now suppose 30 people did this experiment. What do you think the numbers of reds will most likely go from?

From a low of ______ to a high of _______.

Why do you think this?

[3] At the same container, suppose that 50 people each pulled out handfuls of 10 candies, wrote down the number of reds, put the candies back and mixed them up again. Of the 50 people, how many of them do you think would get:

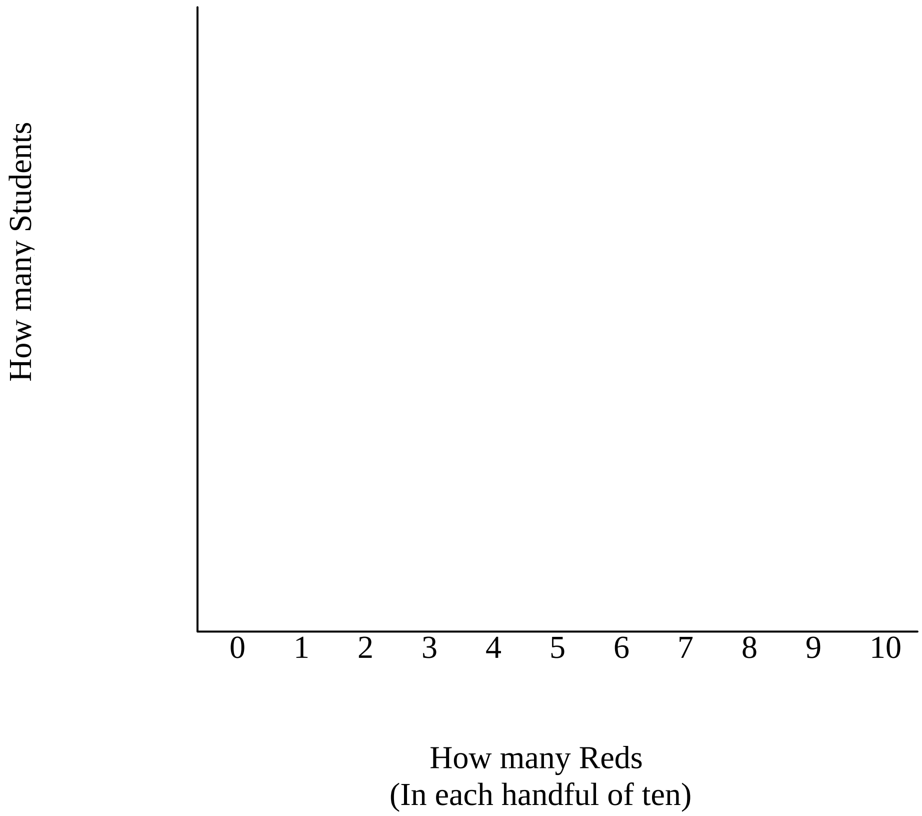
0 Red ? ____
1 Red ? ____
2 Red ? ____
3 Red ? ____
4 Red ? ____
5 Red ? ____
6 Red ? ____
7 Red ? ____
8 Red ? ____
9 Red ? ____
10 Red ? ____

Total : 50 People

Why do you think the numbers you wrote above are reasonable?

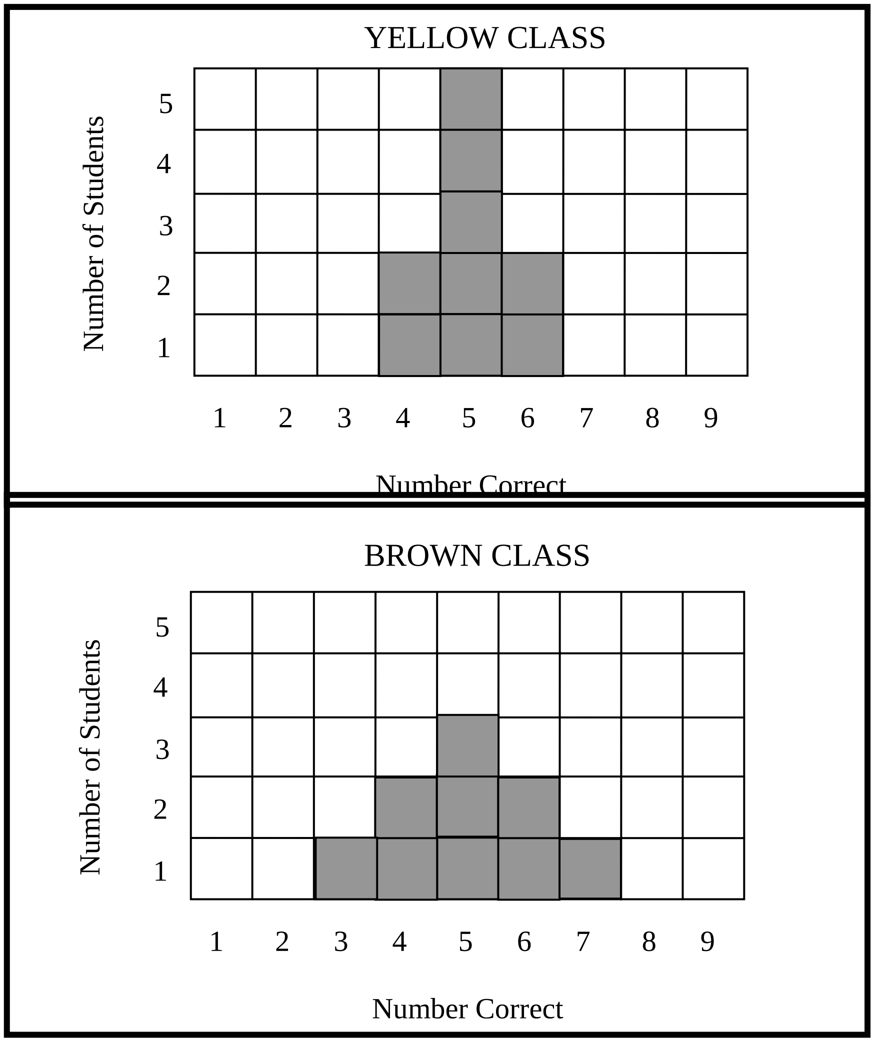
[4] Fifty students lined up at the candy container. Each student pulled a handful of 10 candies, wrote on the chalkboard how many reds they had, and then returned the candies and mixed them all up again.

The class decided to draw a graph of their data. Show below what their graph might look like:



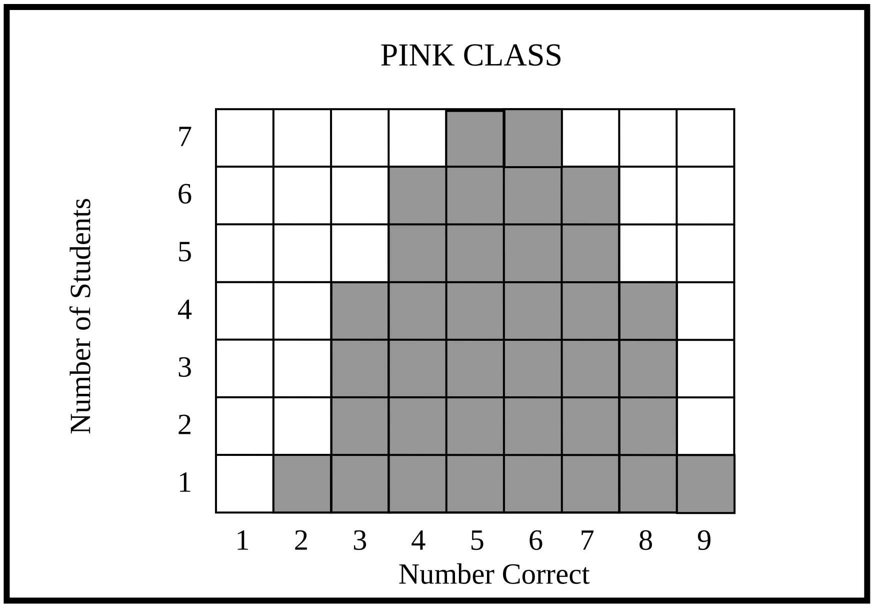
[5] Two schools are comparing some classes to see which school is better at spelling. All the classes took identical tests.

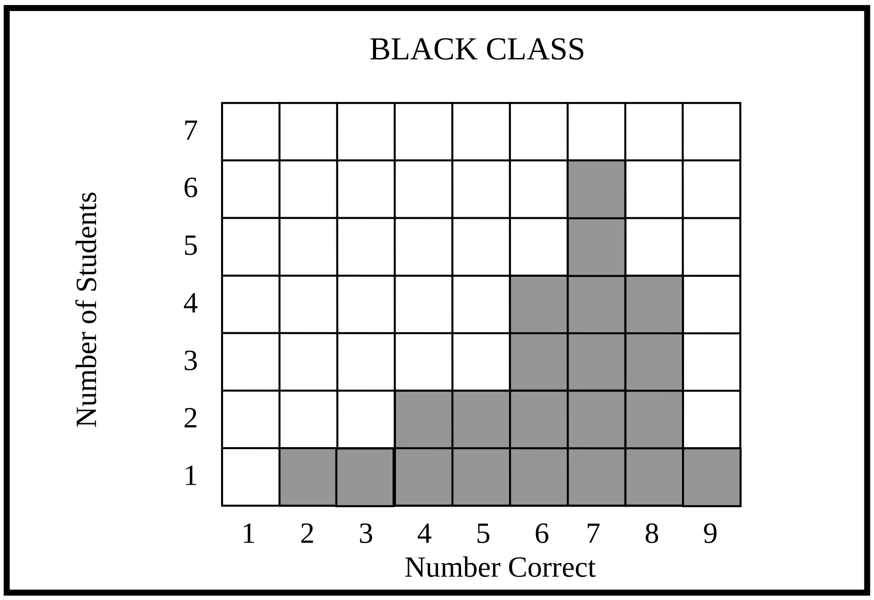
(a) First consider two classes, the YELLOW class and the BROWN class. The scores for the two classes are shown on the two charts below, Each shaded box is one person's test score.



Now look at the scores of all students in each class, and then decide: Did the two classes do equally well on the test, or did one of the classes do better than the other? Explain how you decided.

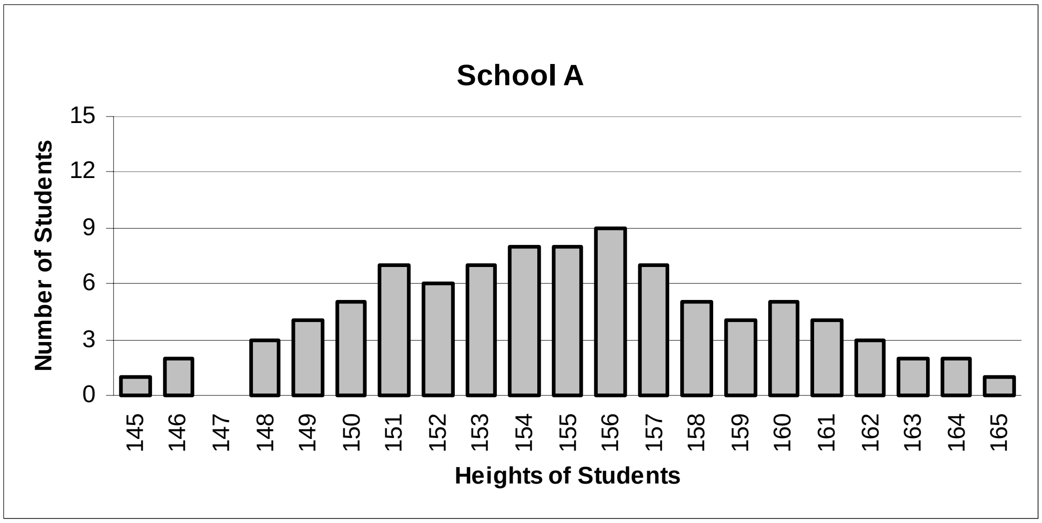
(b) Now consider two more classes, the PINK class and the BLACK class. The scores for the two classes are shown below, and once again each box is one person's test score.

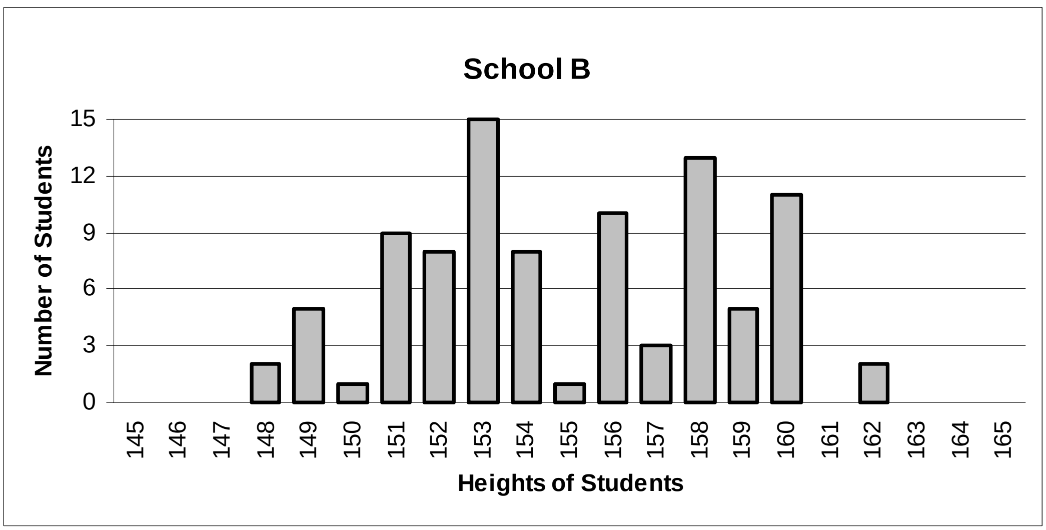




Again look at the scores of all students in each class, and then decide: Did the two classes do equally well on the test, or did one of the classes do better than the other? Explain how you decided.

[6] The following graphs describe some data collected about Grade 7 students' heights (measured in centimeters) in two different schools:





Which graph shows more variability in students' heights?

Explain why you think this.

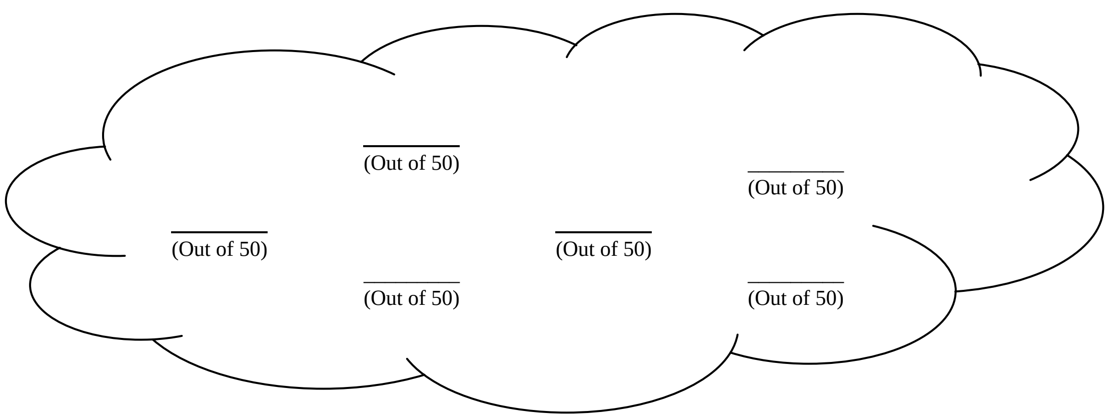
[7] Consider flipping a fair coin.

(a) Mark is curious to see how often the coin lands Heads-up, so he flips it 50 times. How many times out of 50 flips do you think the coin might land Heads-up for Mark?

Why do you think this?

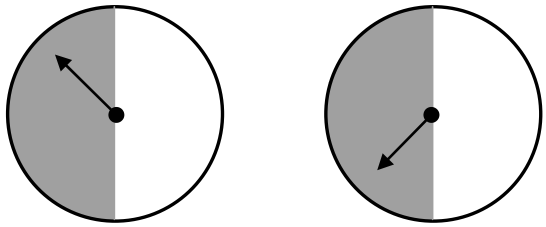
(b) After Mark's first set of 50 flip, he decides to do a second set of 50 flips. How do you think his results on the second set of 50 flips will compare with the results of his first set?

(c) Mark actually has a lot of time on his hands, so the next day he does 6 sets of 50 flips. Write in the numbers for what you think might happen for the number of flips out of 50 the coin would land Heads-up (in each of the 6 sets of 50 flips).



Why did you choose those numbers?

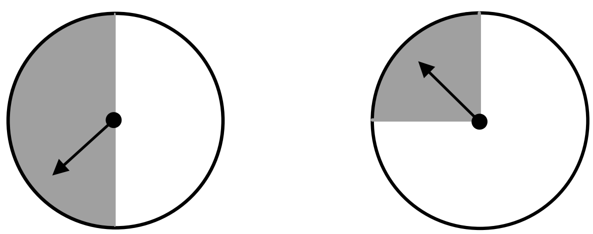
[8] The two fair spinners shown below are a part of a game, which goes like this: A player spins each spinner once, and wins a prize *only if both* arrows land on **black**.



Angela thinks she has a 50-50 chance of winning. Do you agree? ☐ Yes ☐ No

Explain your answer. Why do you think this?

[9] Suppose the game is played with the new spinners shown below. Again, both spinners are spun once, and the arrows must *both* land on **black** in order to win.



What do you think the chances of winning this game would be?

Explain your answer. Why do you think this?

PREINTERVIEW

First Interview Date: Name:

[1] Suppose there is a container with 100 pieces of candy in it. 60 are Red, and 40 are Yellow. The candies are all mixed up in the container.

You reach in and pull out a handful of 10 candies at random.

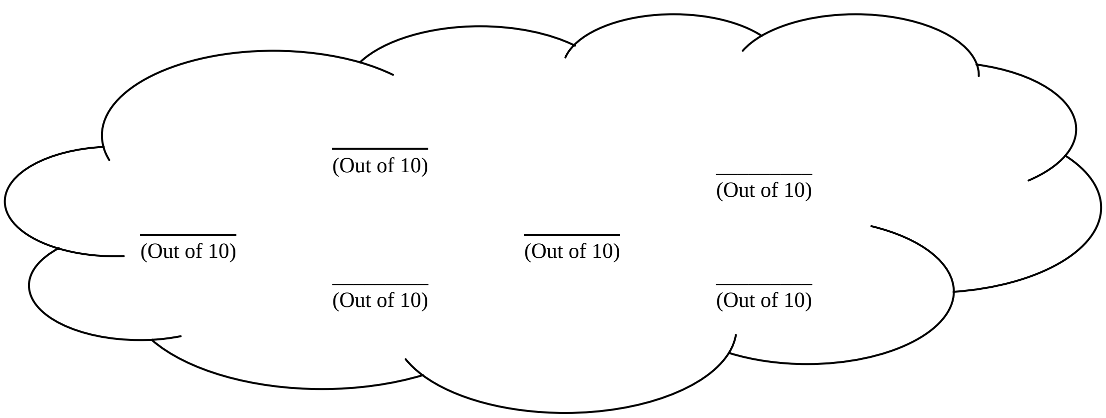
(a) How many red candies do you think you might get?

Why do you think this?

(b) Suppose you do this several times (each time returning the previous handful of 10 candies and remixing the container). Do you think this many reds would come out every time?

Why do you think this?

(c) Suppose six classmates do this experiment (each time returning the previous handful of 10 candies and remixing the container). Write down the number of reds that you think each classmate might get:



Why did you choose those numbers?

[2] Here are some examples of what other people have said for the numbers of reds that they think the six classmates would get in each handful.

i) 7, 9, 7, 6, 8, 7

ii) 6, 7, 5, 8, 5, 4

iii) 6, 6, 6, 6, 6, 6

iv) 2, 5, 4, 3, 6, 4

v) 3, 10, 9, 2, 1, 5

(a) Put a check mark next to any of these that you think might be likely.

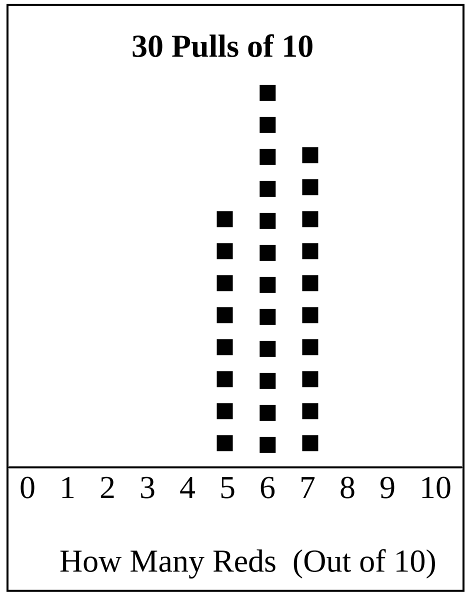
(b) Circle the list that you think best describes what might happen.

(c) Why do think the list you chose best describes what might happen?

[3] Matt took his class to the candy container (100 Candies = 60 Red and 40 Yellow). Then he left the room. When he came back, the class claimed to have pulled 30 samples each of size 10, with replacement. They showed Matt their data and a graph:

Number of Reds
in 30 Samples of 10

7	6	5
5	7	6
6	5	7
7	6	7
6	6	6
7	5	5
5	6	5
6	7	6
7	6	6
5	7	7



Which of the following do you think is *most* likely ? Put a check mark next to it.

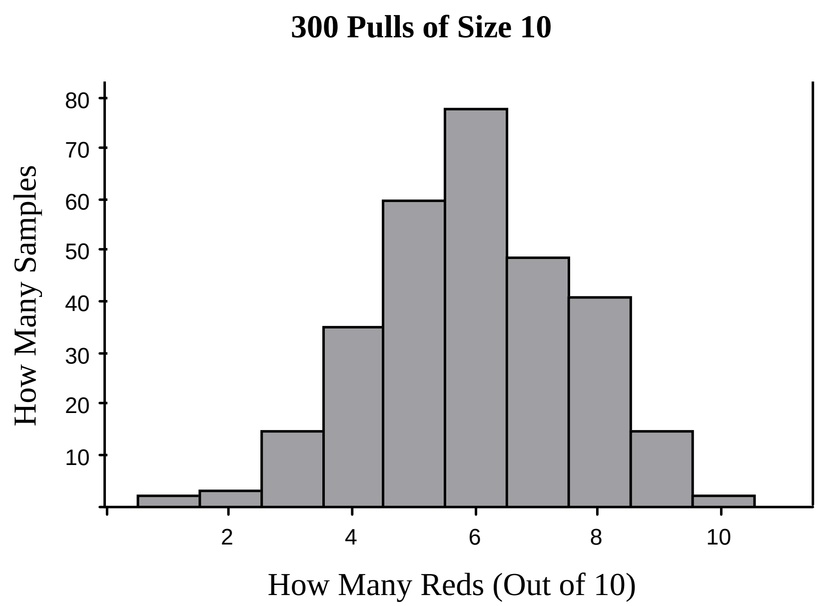
___ Matt’s class just made up these results

___ Those are the actual results of the class samples

___ No one can have much confidence if the results are made up or not.

Explain why you think this is the most likely.

[4] Jen's class also visited the candy container (100 Candies = 60 Red and 40 Yellow). The class claims to have pulled 300 samples each of size 10, with replacement. They showed Jen this graph:



(a) Which of the following do you think is *most* likely ? Put a check mark next to it.

___ Jen's class just made up these results

___ Those are the actual results of the class samples

___ No one can have much confidence if the results are made up or not.

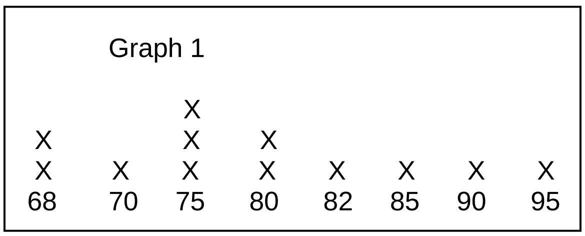
Explain why you think this is the most likely.

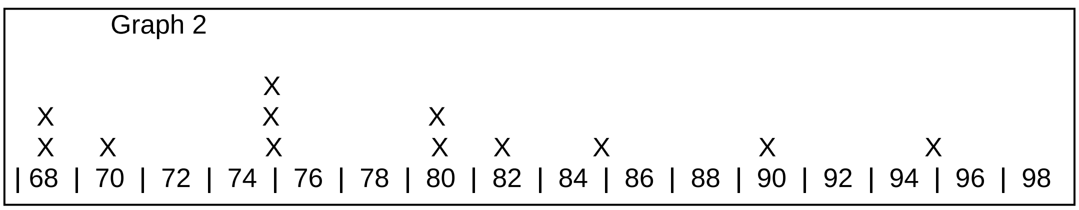
(b) How does the shape of the graph for Jen's class compare to the shape of the graph for Matt's class?

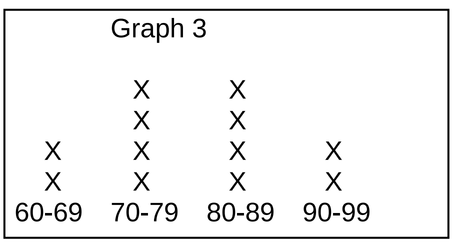
[5] A new car was being tested to see how well the brakes worked. The test engineer measured how many inches the car took to slow from 40 mph to 0 mph; the fewer inches taken, the better the braking power. Twelve trials were run, under the same road conditions and with the same test driver. Here were the results (to the nearest inch):

Stopping Distance (in.)			
68	68	70	75
75	75	80	80
82	85	90	95

The engineer was then trying to decide how to graph the results. She came up with the following three graphs for representing the data:







(a) Do these graphs differ in the way they show the braking power? If so, how?

(b) Do you think one graph shows more variability in the results than the others? Explain.

(c) If the engineer wanted to suggest that the car was fairly consistent in its braking power, which graph would you suggest she use, and why?

[6] A class of twenty-one 6th-grade students wanted to find out some information about MAX train rides. Their first goal was to find out the duration of a ride from Washington Park to Gresham. They all got on the same train, but they sat separately and kept track of the time on their own. Later in class, they were surprised to find that they did not have the same results:

Duration of Ride (Min:Sec , to the nearest second)

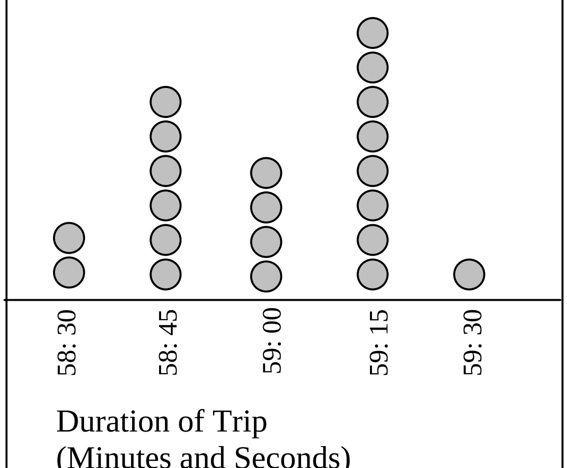
58: 36	58 :36	58: 40	58: 44	58: 51	58: 50
58: 49	58: 50	58 :56	59: 01	59: 02	59: 06
59: 11	59: 09	59: 16	59 :14	59: 15	59: 19
59: 21	59: 20	59: 24			

What are some possible reasons for why the class did not all get the same result?

[7] The class was deciding how to display their data. In Graph 1, they rounded to the nearest 15 seconds. In Graph 2, they rounded to the nearest 5 seconds.

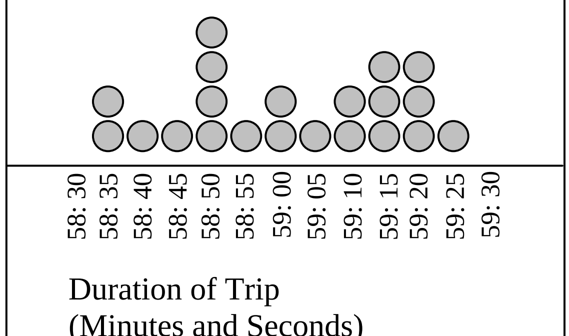
***GRAPH 1** (Rounded to nearest 15 sec.)*

Mean = 59:00
Median = 59:00
Mode = 59:15



***GRAPH 2** (Rounded to nearest 5 sec.)*

Mean = 59:01
Median = 59:00
Mode = 58:50



(a) How do these graphs differ in the stories they tell about the duration of the trip?

(b) Some members of the class argue that the trip was really under 59 minutes, while some argue that it was over 59 minutes. Others claim it was exactly 59 minutes. What do you think about the true duration of the trip, and why do you think this?

(c) Does one graph help you more than the other in making your conclusion?

[8] The Wait-Time for the MAX is defined as the interval of time which starts when one train leaves and ends when the next train arrives. In other words, the Wait-Time is how long there's no train at the station.

A class of twenty students wanted to find out if there was a difference in Wait-Times between Westbound and Eastbound MAX trains. They went and got the following ten Wait-Times for different Westbound trains and ten Wait-Times for different Eastbound trains (rounded to the nearest half-minute):

Data: (Wait-Times in Minutes)

Westbound

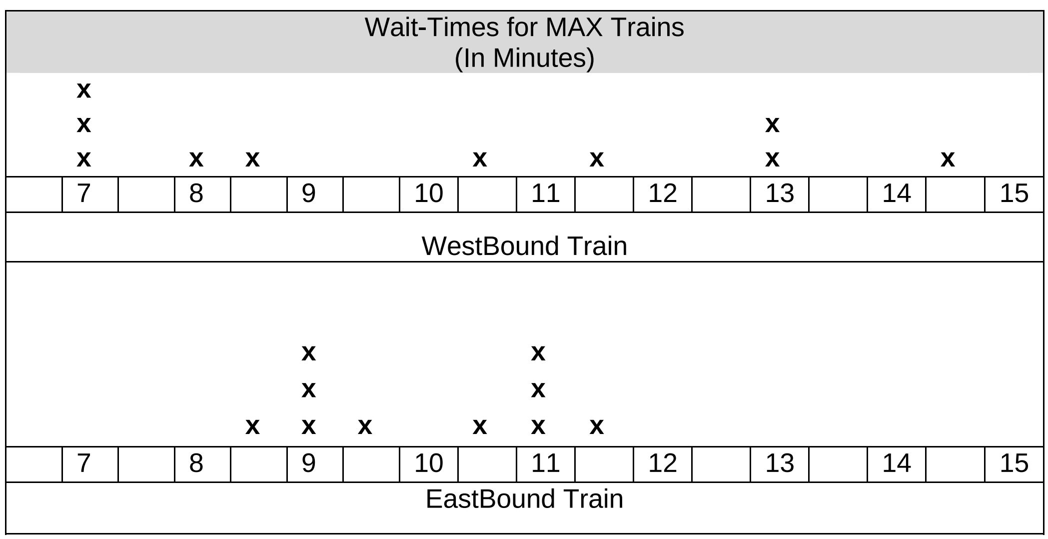
7.0	7.0	7.0	11.5	10.5
8.5	8.0	13.0	14.5.	13.0

Mean = 10.0 min.
Median = 10.5 min.

Eastbound

8.5	9.0	9.0	11.0	11.0
9.5	9.0	11.0	10.5	11.5

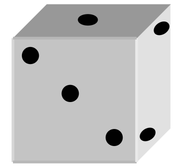
Mean = 10.0 min.
Median = 10.5 min.



(a) What can you conclude about the Wait-Times for the two trains?

(b) One student in class argues that there's really no difference in the Wait-Times of the two trains, since the averages are the same. Do you agree?

[9] Consider a regular, fair, six-sided die. Imagine that you threw the die 60 times. Fill in the table below to show how many times you think each number might come up.



Number that shows on the tossed die	How many times it might come up
1	
2	
3	
4	
5	
6	
Total =	**60**

Why do you think those numbers are reasonable?

[10] For homework, Mr. Blair asked each student in his class to toss a die 60 times and keep track of how many times each of the 6 sides came up. Below are the results turned in the next day by four students (Riki, for example, reported that Side 1 came up 7 times in 60 tosses).

	Riki	**Lynn**	**Lee**	**Pat**
Side that came up				
1	7	10	10	2
2	12	11	10	15
3	6	10	10	10
4	9	10	10	28
5	14	9	10	1
6	12	10	10	4

Only one of these students actually rolled the die. The other three students just made up their results before class. What do you think is most likely?

___i) Riki really rolled it

___ii) Lynn really rolled it

___iii) Lee really rolled it

___iv) Pat really rolled it

___v) No one can say. Any of the 4 students is equally likely to have really rolled it.

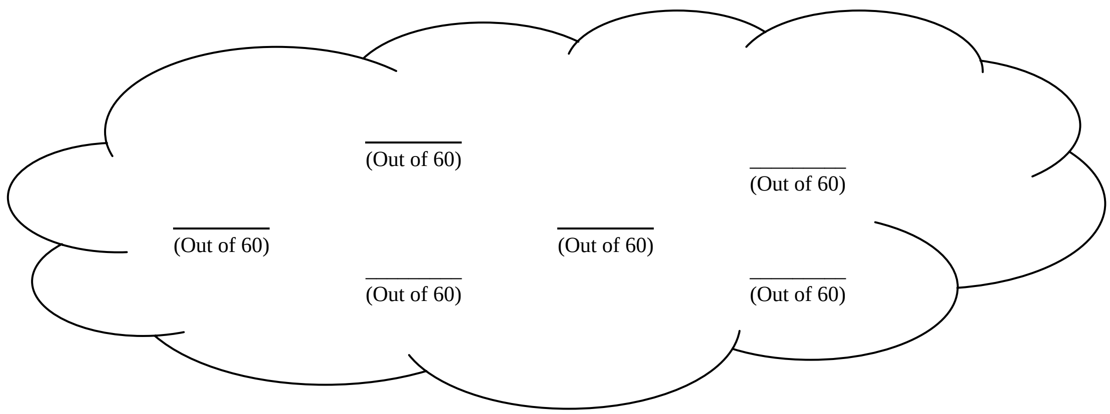
Explain your reasoning.

[11] Look back at Question [9] to see how many "5"s you predicted in 60 tosses.

(a) If you did another set of 60 tosses, do you think you would get that many "5"s again?

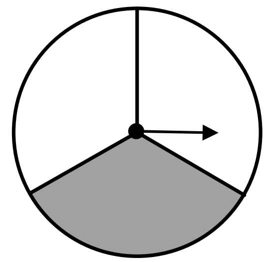
Why or why not?

(b) If six friends took turns tossing the die 60 times each, write down how many "5"s you think each friend might get in their 60 tosses:



Why did you choose those numbers?

[12] The spinner at the right has three regions of equal area: Two of the three regions are White and one of the three regions is Black.



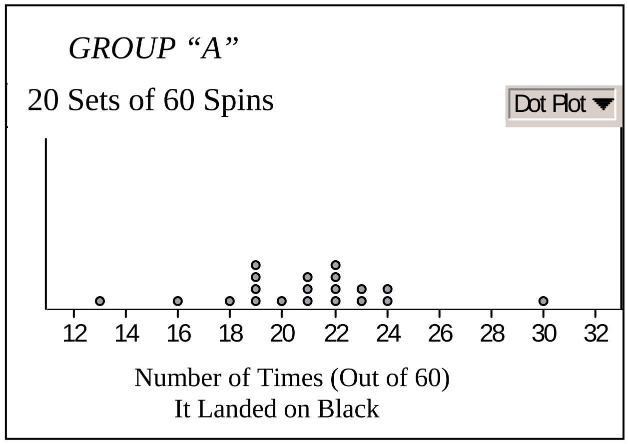
(a) If you spun this 3 times, would you be surprised if you got more Black than White?

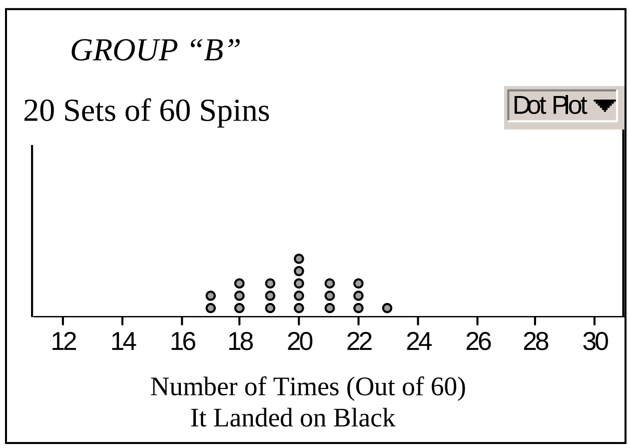
(b) If you spun this 12 times, would you be surprised if you got more Black than White?

(c) If you spun this 60 times, would you be surprised if it landed on White 30 times?

[13] Ron split his class into two groups, and he told each group to conduct the following experiment twenty times: Spin the spinner 60 times, and write down how many times it landed on Black. So, each group was supposed to do twenty sets of 60 spins.

Ron went for some coffee, and when he returned he saw these two graphs:





(a) What are some similarities and differences you notice in the two graphs above?

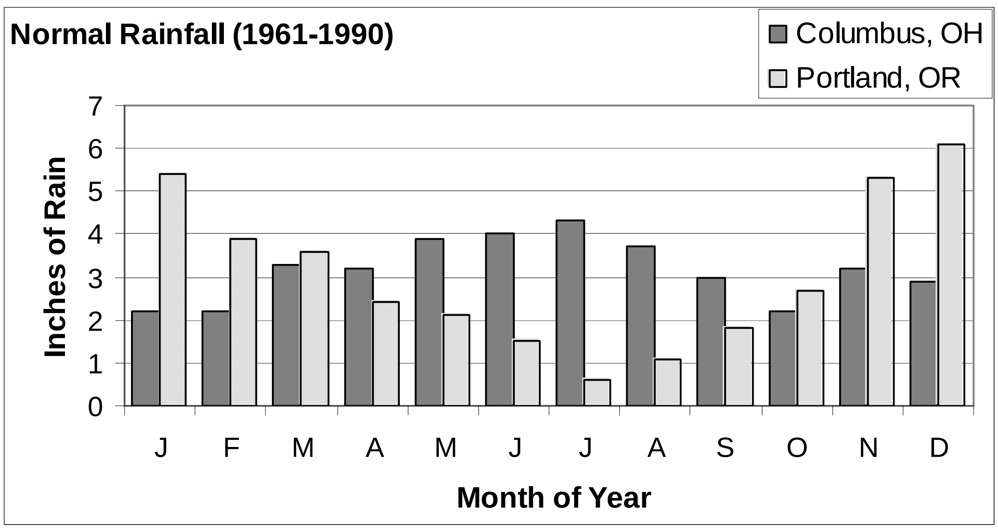
(b) Does one graph or the other look more like what you would have expected? Explain.

POSTSURVEY (DATA & GRAPHS)

Mth 212 Written Reflection: Data & Graphs Name: __

[1] The following data was taken from a National Weather Service. They kept records of the rainfall in cities to see how much rain fell each month. After 30 years, they averaged the amounts of rainfall in each month: This is called the average, or Normal Rainfall for the 30-year period.

(a) In the bar chart below, the normal monthly rainfall data for both Portland (Oregon) and Columbus (Ohio) are graphed together:



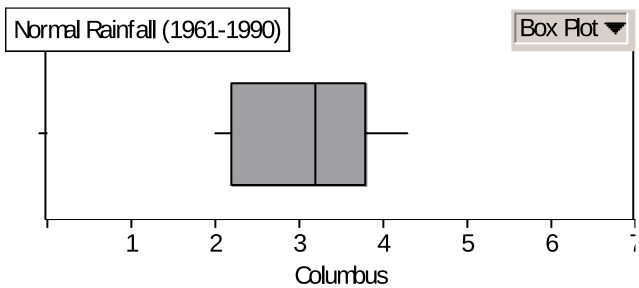
Columbus
Mean = 3.18 in.
Median = 3.20 in.

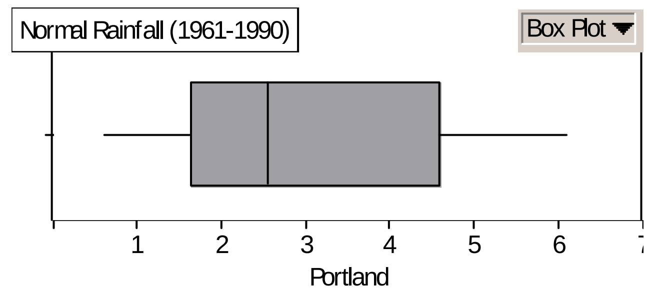
Portland
Mean = 3.04 in.
Median = 2.55

i) What do you think are some causes for the different patterns of rain in the two cities?

ii) Adam and Zain are two Math 212 students who were discussing the data. Adam said Portland was rainier because it got the highest amount of rain a month. What do you think Adam is thinking when he says this?

(b) Here are two boxplots that show the same data for the normal monthly rainfall in both Columbus and Portland:

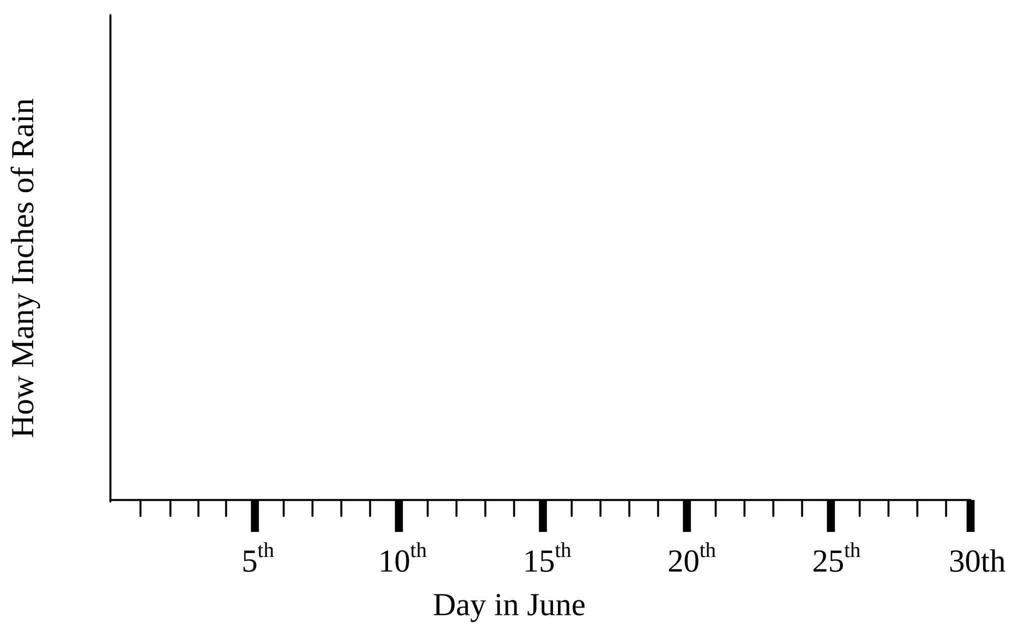




i) Zain said Columbus was rainier because the average monthly rainfall was higher than Portland. What do you think Zain is thinking when he says this?

ii) Which city do *you* think is rainier, and why?

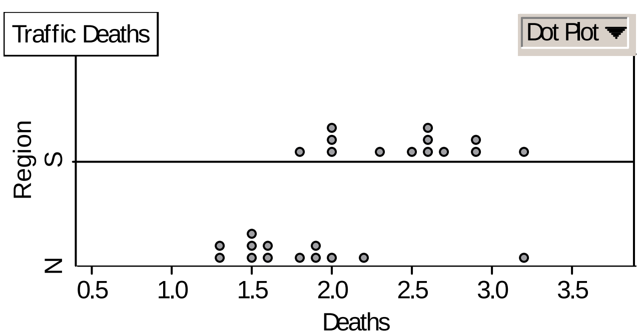
c) In Columbus, the normal monthly rainfall for the month of June is reported as 4 inches. Draw a graph below which shows how many inches of rain Columbus might get for each day in June (assuming that the average rainfall for the entire month is 4 inches).

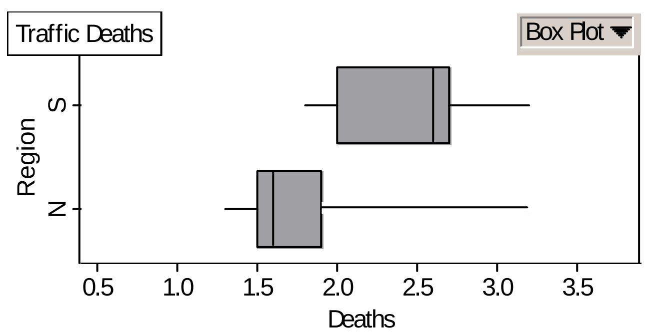


[2] The National Safety Council groups the following 25 States plus the District of Columbia into two regions: South and Northeast. The following data and graphs show the number of traffic deaths in a recent year per 100 million vehicle miles driven:

Motor Vehicle Traffic Deaths per 100 Million Vehicle Miles Driven

South		**Northeast**	
Alabama	2.6	Connecticut	1.5
Arkansas	2.9	Delaware	2.2
Florida	2.7	District of Columbia	1.6
Georgia	2.0	Maine	1.8
Kentucky	2.6	Maryland	1.9
Louisiana	2.5	Massachusetts	1.3
Mississippi	3.2	New Hampshire	1.6
North Carolina	2.3	New Jersey	1.5
Oklahoma	2.0	New York	2.0
South Carolina	2.9	Pennsylvania	1.9
Tennessee	2.6	Rhode Island	1.3
Texas	2.0	Vermont	1.5
Virginia	1.8	West Virginia	3.2





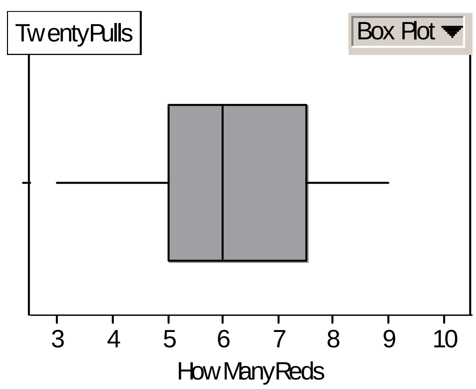
(a) How do the traffic deaths rates in the South compare with those in the Northeast ?

(b) What factors do you think might help to explain the difference between the South and the Northeast ?

POSTSURVEY (SAMPLING)

Mth 212 Written Reflection: Sampling Name:

[1] Consider the container of candy that holds 100 pieces (60 Red and 40 Yellow). Suppose that 20 people lined up at the container. Each person pulled out a handful of 10 candies at a time , wrote down the number of reds, put the candies back and mixed them up again. They made a boxplot of their results, and the graph looked like this:



Median = 6
Minimum = 3
Maximum = 9

a) When we look at the boxplot, we only get a bit of information about the entire data set. Write down what you think all 20 results might have been (for the number of red in each handful of ten):

____ ____ ____ ____ ____ ____ ____ ____ ____ ____

____ ____ ____ ____ ____ ____ ____ ____ ____ ____

b) Mike was surprised that nobody got 0 or 1 Red candy in their handful. He decides that he's going to try it with more than 20 people.

How many people do you think Mike should have do this so that at least one person gets 0 or 1 Red candy in their handful ?

How did you decide on that answer?

[2] Now suppose there is a LARGER container with 1000 pieces of candy in it. 600 are Red, 400 are Yellow. The candies are all mixed up in the container.

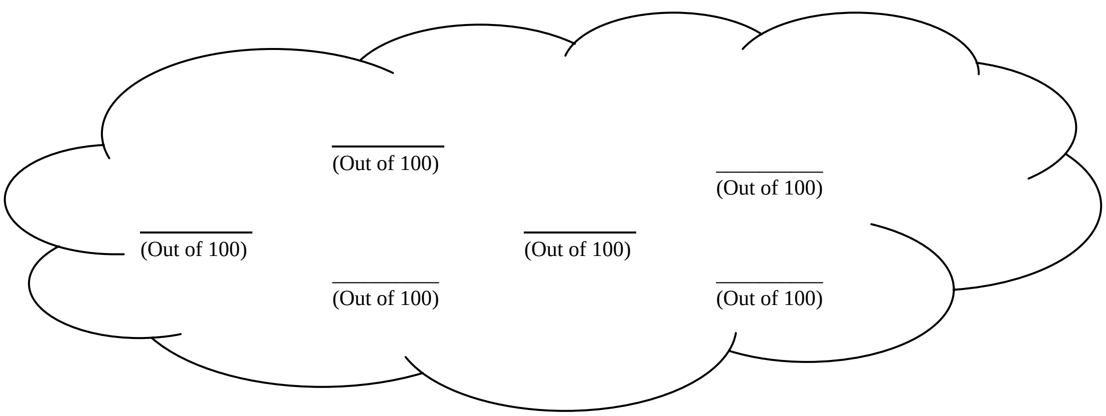
You reach in and pull out a handful of 100 candies at random.

(a) How many red candies do you think you will get?

(b) Suppose you do this several times (each time returning the previous handful of 100 candies and remixing the container). Do you think this many reds would come out every time?

Why do you think this?

(c) Suppose six classmates do this experiment (each time returning the previous handful of 100 candies and remixing the container). Write down the number of reds that you think each classmate obtained:



Why did you choose those numbers?

[3] Suppose 30 people did this experiment – pulled one hundred, candies from the LARGE container (600 Red and 400 Yellow), wrote down the number of reds, then returned the hundred and remixed all the candies.

What do you think the numbers of reds will most likely go from?

From a low of ______ (out of 100) to a high of _______ (out of 100).

Now suppose 300 people did this experiment. What do you think the numbers of reds will most likely go from?

From a low of ______ (out of 100) to a high of _______ (out of 100).

Why do you think this?

[4] Suppose that 50 people each pulled out 100 candies from the LARGE container (600 Red and 400 Yellow), wrote down the number of reds they pulled, put the candies back and mixed them up again. Of the 50 people, how many of them do you think would get:

0 -10 Red ? ____
11-20 Red ? ____
21-30 Red ? ____
31-40 Red ? ____
41-50 Red ? ____
51-60 Red ? ____
61-70 Red ? ____
71-80 Red ? ____
81-90 Red ? ____
91-100 Red ? ____

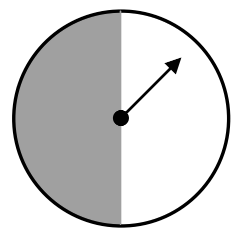
Total : 50 People

Why do you think the numbers you wrote above are reasonable?

POSTSURVEY (PROBABILITY)

Mth 212 Written Reflection: Probability Name: ______________

[1] Consider thespinner on the right:

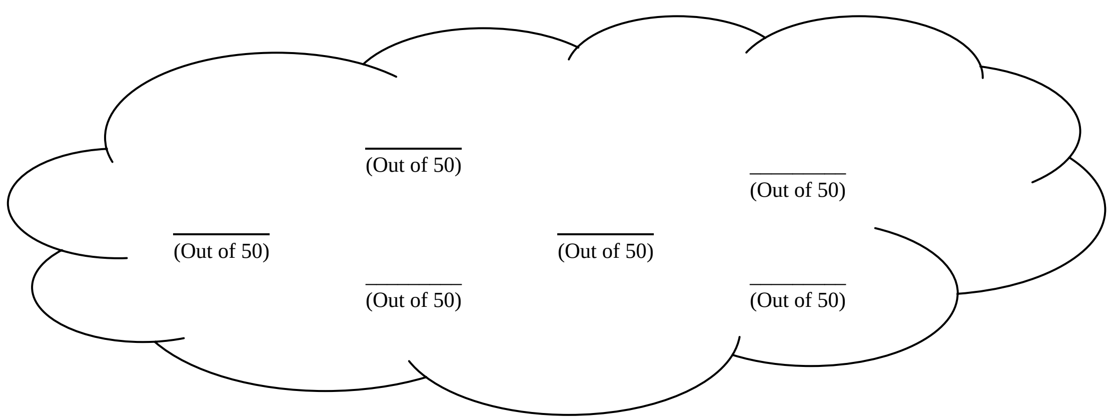


(a) Matt is curious to see how often the spinner lands on black, so he spins it 50 times. How many times (out of 50 tries) do you think the arrow might land black?

Why do you think this?

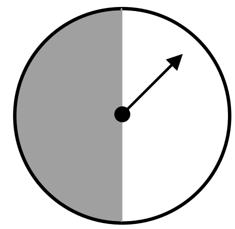
(b) After Matt's first set of 50 spins, he decides to do a second set of 50 spins. How do you think his results on the second set of 50 spins will compare with the results of his first set?

(c) Matt actually has a lot of time on his hands, so the next day he does 6 sets of 50 spins. Write a list that would describe what you think might happen for the number of spins out of 50 the spinner would land on black in each of the 6 sets of 50 spins.



Why did you choose those numbers?

[2] Suppose 30 students did this experiment – Each student spun the spinner 50 times and wrote down how many times (out of 50 tries) the arrow landed on black.



What do you think the students' numbers (of times the arrow lands on black) will most likely go from?

From a low of ______ to a high of _______.

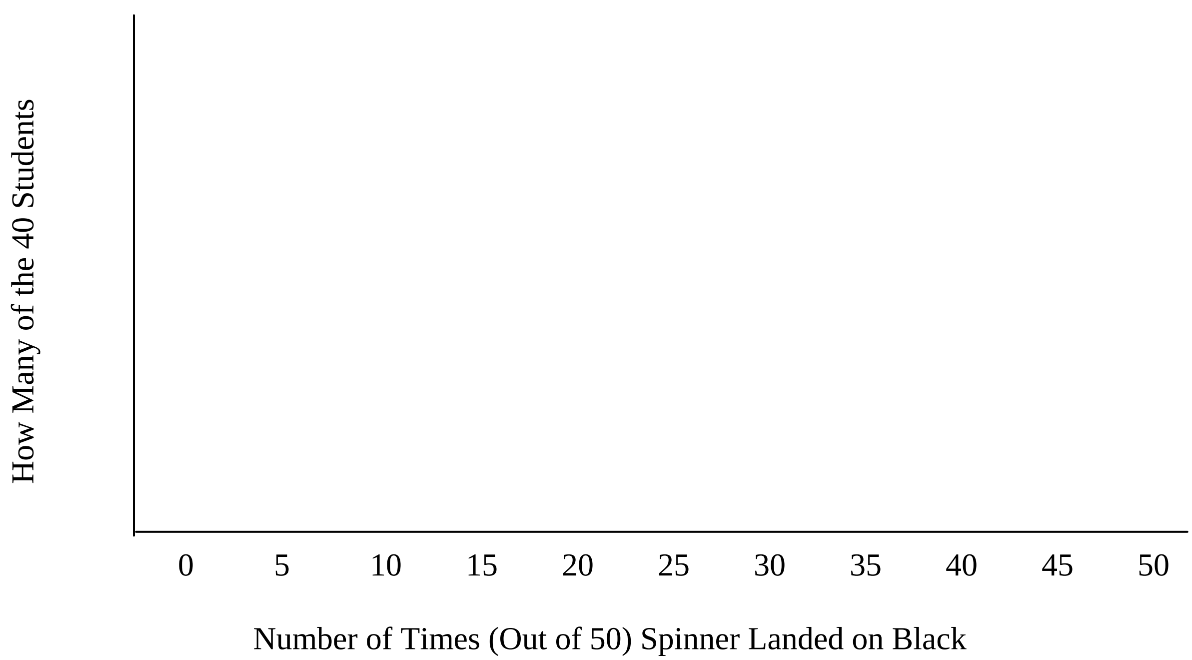
Now suppose 300 students did this experiment. What do you think the numbers (of times the arrow lands on black) will most likely go from?

From a low of ______ to a high of _______.

Why do you think this?

[3] Forty students lined up at the spinner. Each student spun it 50 times and wrote down how many times (out of 50 tries) the arrow landed on the shaded part.

At the end of class, they decided to make a graph of their data. Show below what their graph might look like:



POSTINTERVIEW

Second Interview Date: Name:

[1] Suppose there is a large container with 1000 pieces of candy in it. 600 are Red, 400 are Yellow. The candies are all mixed up in the container.

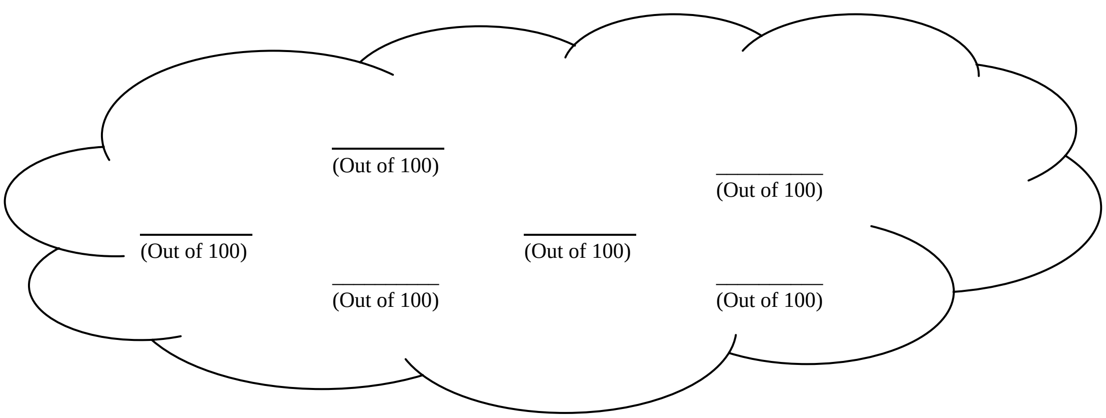
You reach in and pull out a handful of 100 candies at random.

(a) How many red candies do you think you will get?

(b) Suppose you do this several times (each time returning the previous handful of 100 candies and remixing the container). Do you think this many reds would come out every time?

Why do you think this?

(c) Suppose six classmates do this experiment (each time returning the previous handful of 100 candies and remixing the container). Write down the number of reds that you think each classmate obtained:



Why did you choose those numbers?

[2] Here are some examples of what other people have said for the numbers of reds that they think the six classmates would get in each handful. Put a check mark next to any of these that you think might be likely:

i) 72, 91, 74, 63, 81, 78

ii) 61, 73, 56, 69, 59, 48

iii) 60, 60, 60, 60, 60, 60

iv) 53, 41, 34, 60, 46, 52

v) 61, 66, 62, 62, 60, 59

vi) 30, 10, 90, 20, 60, 50

(a) Put a check mark next to any of these that you think might be likely.

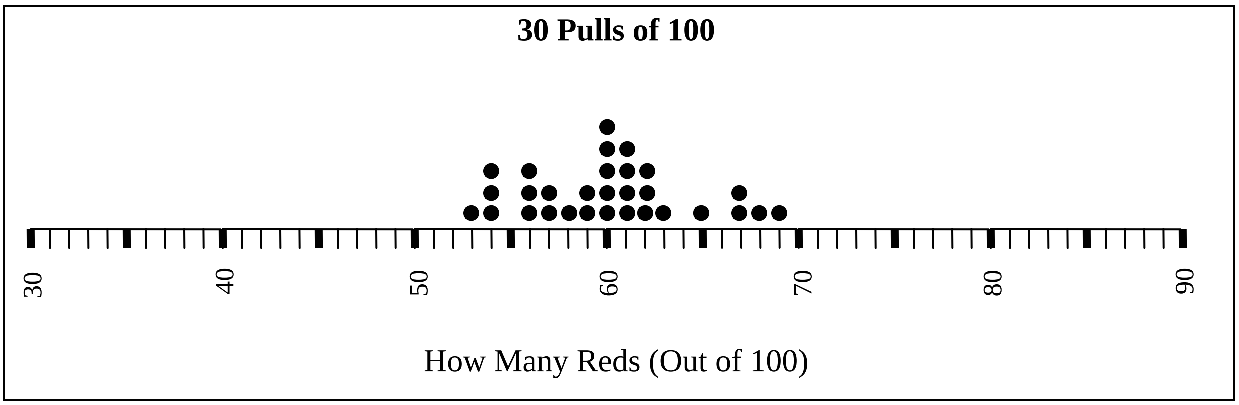
(b) Circle the list that you think best describes what might happen.

(c) Why do think the list you chose best describes what might happen?

[3] Craig took his class to a large candy container (1000 Candies = 600 Red and 400 Yellow). Then he left the room. When he came back, the class claimed to have pulled 30 samples each of size 100, with replacement. They showed Craig their data and a graph:

Number of Reds in 30 Samples of Size 100

53	54	54	54	56	56
56	57	57	58	59	59
60	60	60	60	60	61
61	61	61	62	62	62
63	65	67	67	68	69



Which of the following do you think is *most* likely ? Put a check mark next to it.

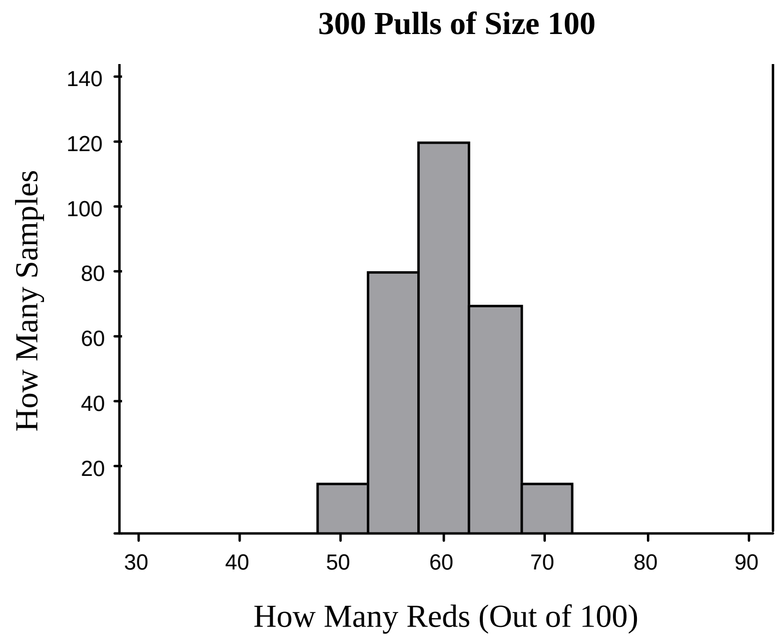
___ Craig's class just made up these results

___ Those are the actual results of the class samples

___ No one can have much confidence if the results are made up or not.

Explain why you think this is the most likely.

[4] Marj also took her class to a large container (1000 Candies = 600 Red and 400 Yellow). Then she left the room. When she came back, the class claimed to have pulled 300 samples each of size 100, with replacement. They showed Marj this graph:



(a) Which of the following do you think is *most* likely ? Put a check mark next to it.

___ Marj's class just made up these results

___ Those are the actual results of the class samples

___ No one can have much confidence if the results are made up or not.

Explain why you think this is the most likely.

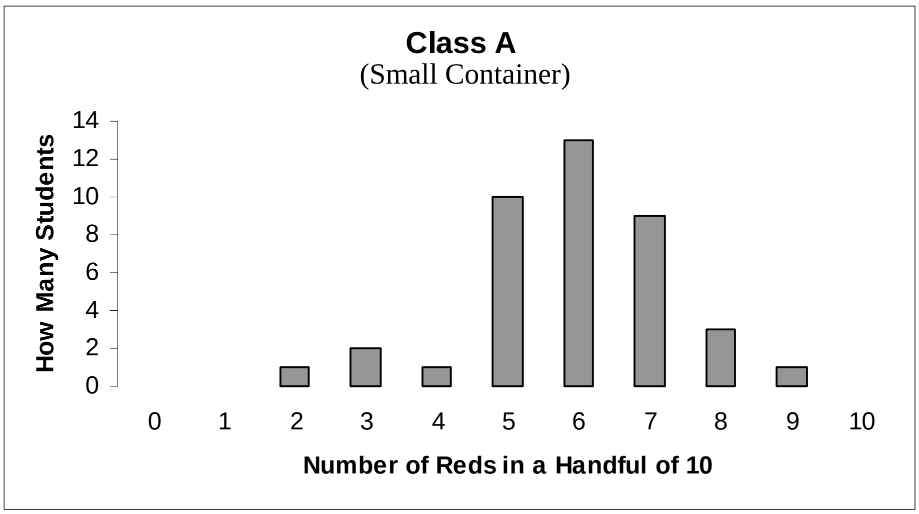
(b) How does the shape of the graph for Marj's class compare to the shape of the graph for Craig's class?

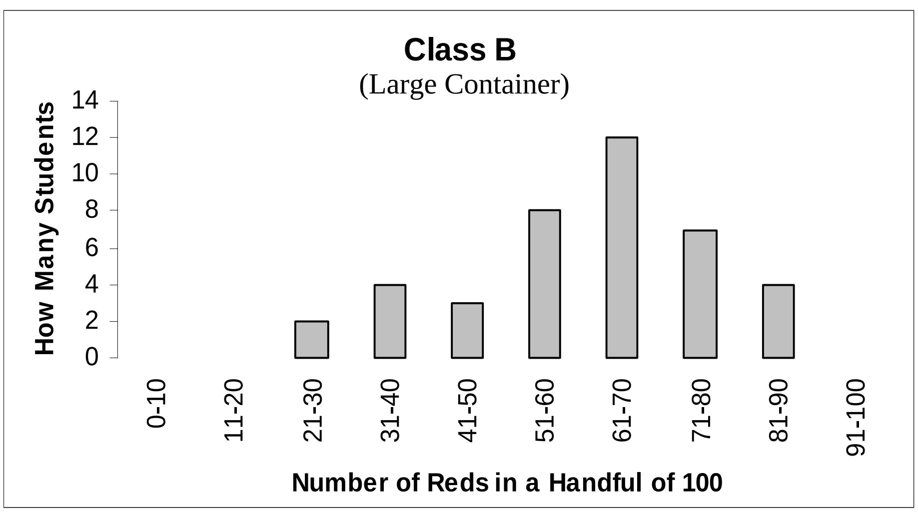
[5] On a day of planned absence from school, Keith left these instructions for his two classes:

- He told the forty students in "Class A" to go to the SMALL container (100 Candies = 60 Red and 40 Yellow). They were each supposed to draw small handfuls of 10 candies (with replacement after each draw of 10).

- He told the forty students in "Class B" to go to the LARGE container (1000 Candies = 600 Red and 400 Yellow). They were each supposed to draw small handfuls of 100 candies (with replacement after each draw of 100).

When Keith came back the next day, he saw these graphs and sets of data for the two classes:





Keith suspects that one of the two classes just made up the data and didn't really carry out the experiment. What do you think? That is, based on the two graphs shown above, do think one graph is likelier than the other to reflect made-up data ?

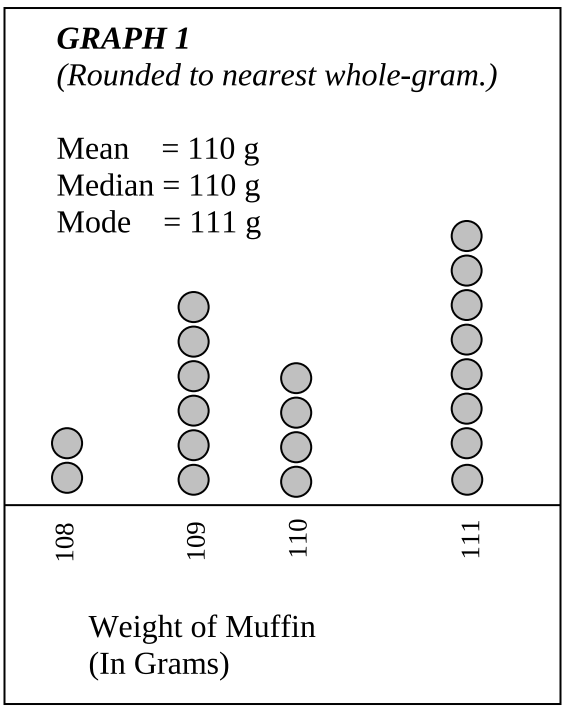
[6] A class of twenty 6th-grade students wanted to find out the muffins weighed from some local bakeries. Their first goal was to see how good their measurement skills were, so they decided to find out the weight of a single muffin from the East End Bakery. They purchased one muffin, and took turns weighing it. They were surprised to find that they did not have the same results:

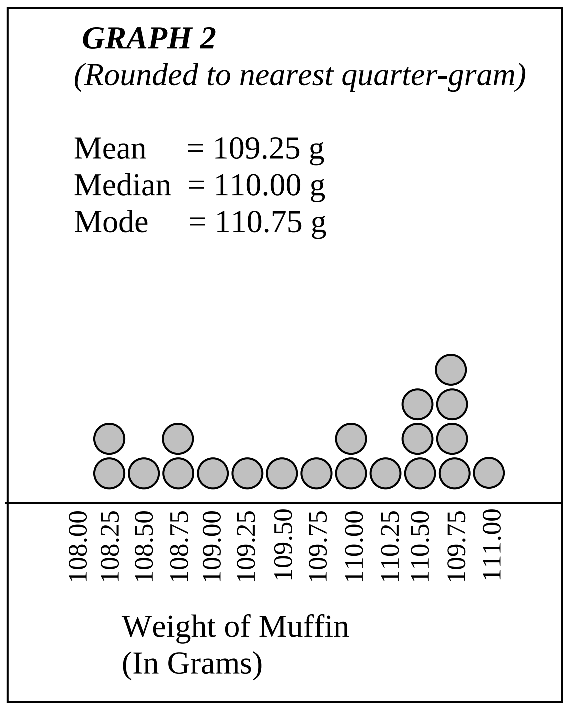
Weight of One Muffin (To Nearest 1/100 of a Gram)

108.23	108.24	108.51	108.75	108.74
108.98	109.24	109.49	109.76	109.98
110.02	110.20	110.50	110.52	110.53
110.75	110.76	110.78	110.74	111.00

What are some possible reasons for why the class did not all get the same result?

[7] The class was deciding how to display their data. In Graph 1, they rounded to the nearest whole gram. In Graph 2, they rounded to the nearest ¼ - gram.





(a) How do these graphs differ in the stories they tell about the weight of the muffin ?

(b) Some members of the class argue that the muffin was really under 110 grams, while some argue that it was over 110 grams. Others claim it was exactly 110 grams. What do you think about the true weight of the muffin, and why do you think this?

(c) Does one graph help you more than the other in making your conclusion?

[8] A different class of 35 students wanted to know how much a “Grande Muffin” from the West End Bakery weighed.. They decided that each student should buy one “Grande Muffin” at different times during the week, and weigh the muffins before eating them.

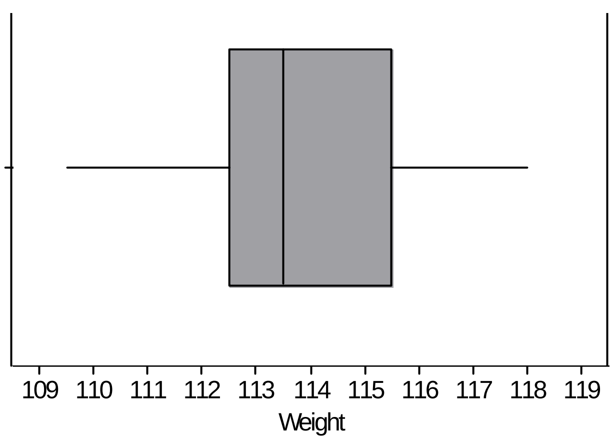
They recorded their 35 muffin weights (rounded to the nearest half-gram, see below), and summarized the results with a boxplot and a histogram:

Data: (Weight of Muffin)

109.5	109.5	110.0	111.0	111.5
112.0	112.0	112.5	112.5	112.5
113.0	113.0	113.0	113.0	113.0
113.0	113.0	113.5	114.0	114.0
114.5	114.5	114.5	114.5	115.0
115.0	115.0	115.5	115.5	116.5
117.0	117.0	117.0	117.5	118.0

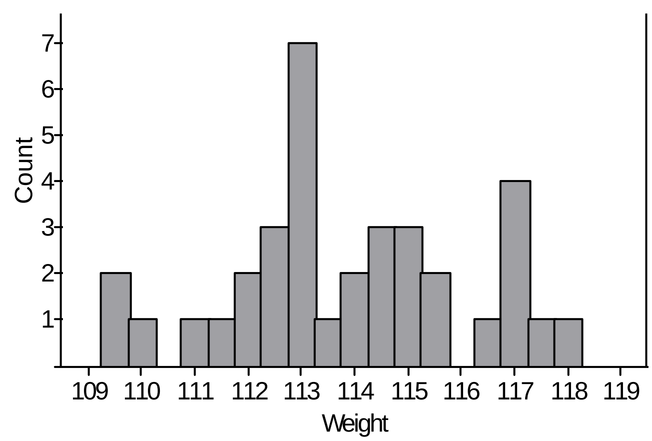
Mean	= 113.79 g
Mode	= 113.00 g
Median	= 113.50 g
Minimum	= 109.50 g
Maximum	= 118.00 g

Boxplot: (35 Muffins)



Weight of Muffin (to nearest half-gram)

Histogram: (35 Muffins)



Weight of Muffin (to nearest half-gram)

a) Suppose you bought a “Grande Muffin” from the West End Bakery. How much do you think your muffin would weigh? Explain your reasoning.

b) What are some similarities and differences in the way these two graphs present the data?

c) Do you think one graph tells you more about the variation in the data than the the other graph? Explain your thinking.

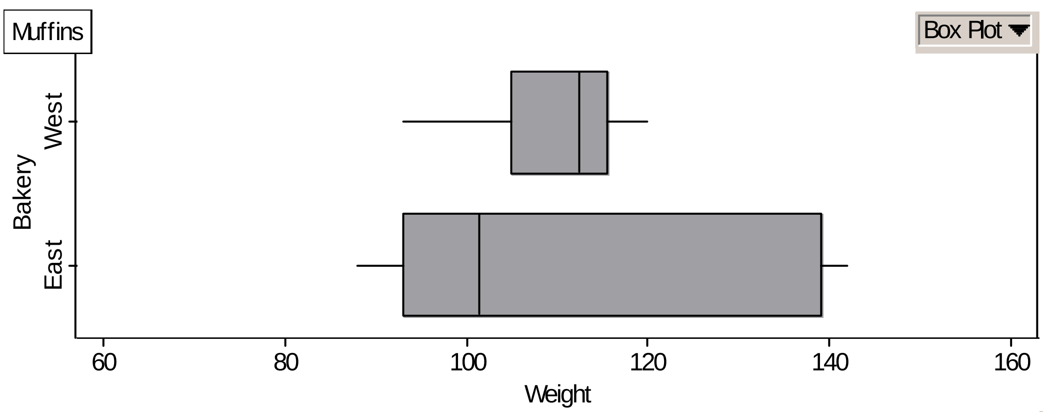
[9] The MathTeam must choose a new bakery to supply them with muffins. The MathTeam wants muffins that are usually at least 110 grams in weight.

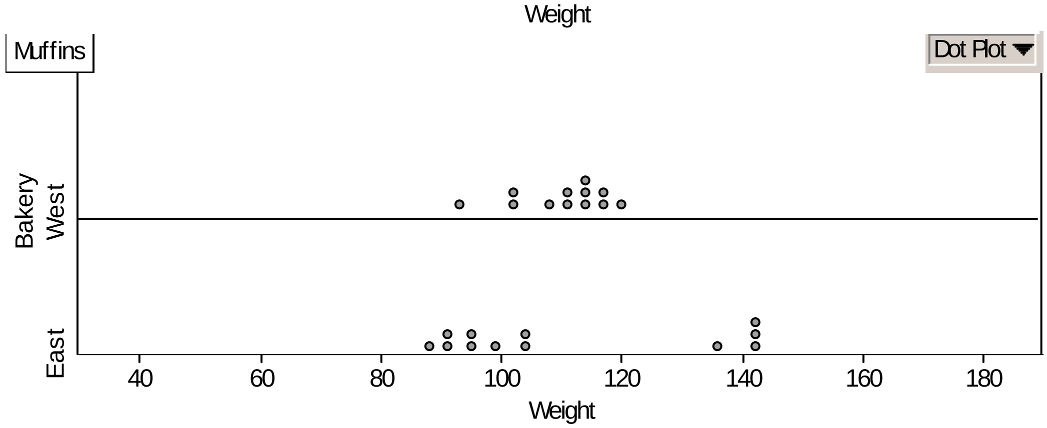
So, the MathTeam samples twelve muffins from the West End Bakery and twelve muffins from the East Side Bakery. The weight of the muffins from the different bakeries were as follows:

West:	93	102	102	108	Mean =	110.25
	111	111	114	114	Median=	112.50
	114	117	117	120	Mode =	114.00

East :	88	91	91	95	Mean =	110.75
	95	99	104	104	Median=	101.50
	136	142	142	142	Mode =	142.00

A Boxplot and Dotplot were used to portray the data:

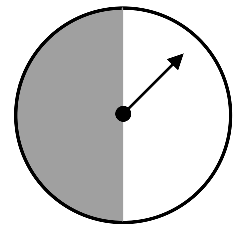




i) Do you think that one bakery would better meet the needs of the MathTeam over the other bakery? Explain your thinking.

ii) Is one graph or the other more helpful in showing the differences between the two bakeries? Explain your thinking.

[10] Consider the spinner on the right:

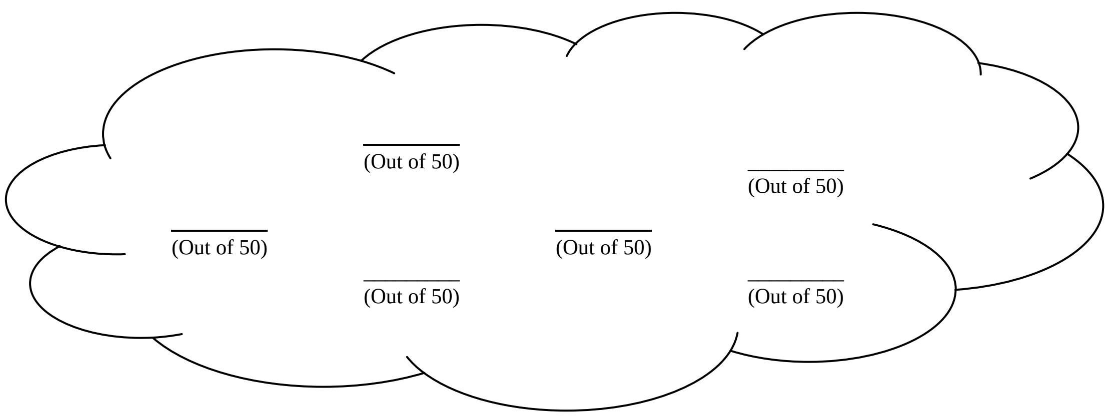


a) Matt is curious to see how often the spinner lands on black, so he spins it 50 times. How many times (out of 50 tries) do you think the arrow might land black?

Why do you think this?

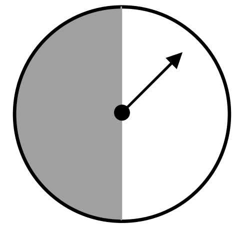
b) After Matt's first set of 50 spins, he decides to do a second set of 50 spins. How do you think his results on the second set of 50 spins will compare with the results of his first set?

c) Matt actually has a lot of time on his hands, so the next day he does 6 sets of 50 spins. Write a list that would describe what you think might happen for the number of spins out of 50 the spinner would land on the shaded part in each of the 6 sets of 50 spins.



Why did you choose those numbers?

[11] Here are some examples of what other people have said for the number of times (out of 50 spins) the arrows would land on black in each of the 6 sets of 50 spins. Put a check mark next to any of these that you think might be likely:



i) 38, 43, 36, 26, 41, 33

ii) 26, 32, 22, 29, 24, 19

iii) 25, 25, 25, 25, 25, 25

iv) 15, 19, 11, 25, 21, 23

v) 24, 25, 26, 25, 24, 26

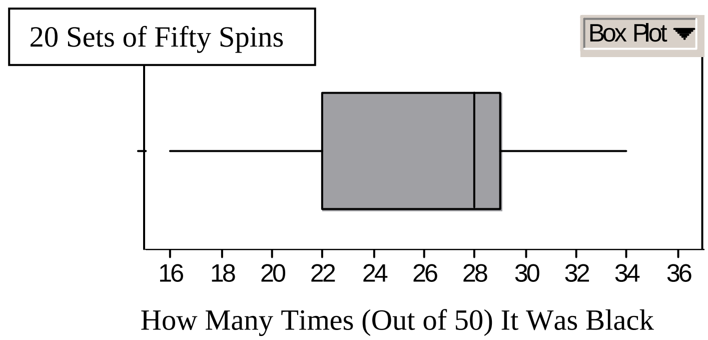
vi) 30, 10, 45, 20, 25, 35

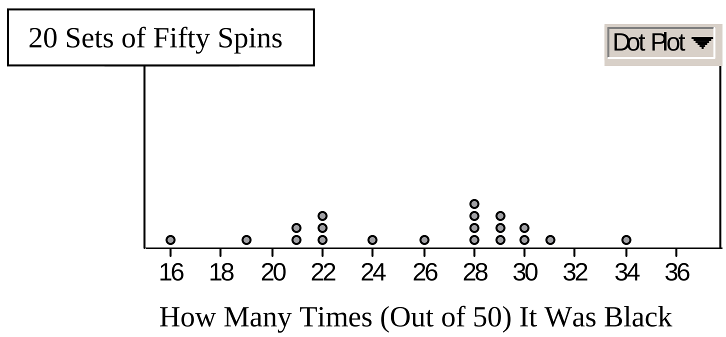
(a) Put a check mark next to any of these that you think might be likely.

(b) Circle the list that you think best describes what might happen.

(c) Why do think the list you chose best describes what might happen?

[12] Twenty students lined up at the spinner. Each student spun it 50 times and wrote down how many times (out of 50 tries) the arrow landed on the shaded part. At the end of class, they decided to make graphs of their data. Their boxplot and dot plot looked like this:





* Keith argued that something was wrong with the experiment because no one got exactly 25 out of 50 spins landing on black.

* Karen argued that, since the mode (and median) was 28 out of 50 spins landing on black, something was wrong with the spinner.

* Jeanette argued that the maximum of 34 out of 50 spins landing on black would not have happened unless something was wrong with the spinner.

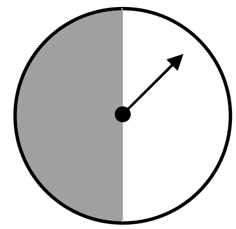
* Marjorie argued there was nothing wrong with the spinner, since she had expected the results to look like this.

(a) What do you think about each person's argument?

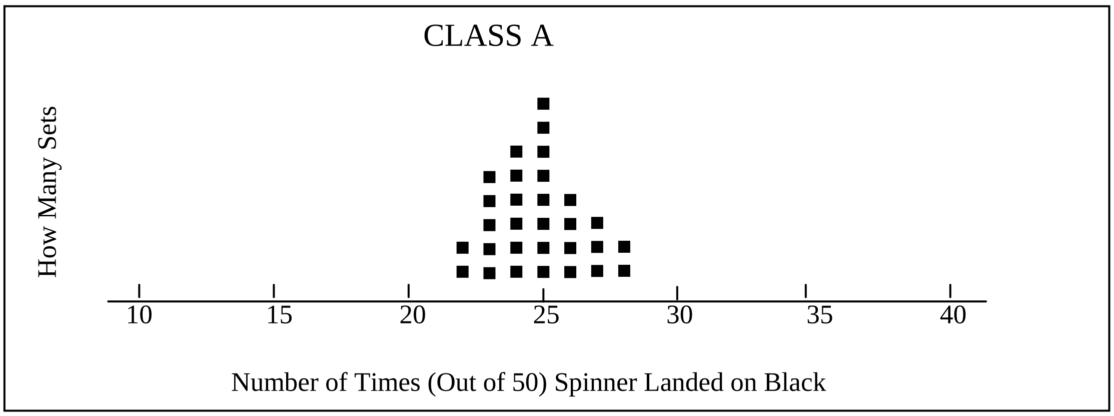
(b) Is there a different argument that you would make?

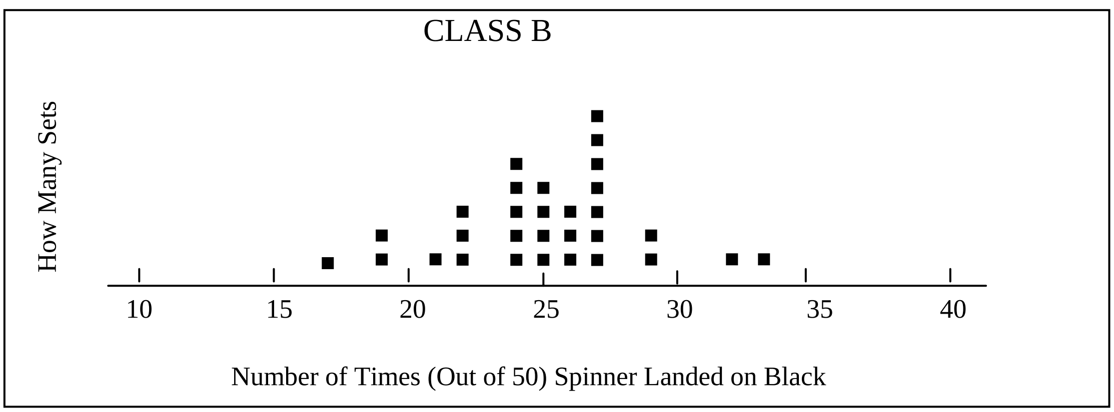
Explain.

[13] On a day of sick leave, Mr. Shaw left instructions for Class A to conduct 30 sets of 50 spins. For each set of 50 spins, they were supposed to record how many times (out of 50) the spinner landed on the shaded part. Then they had to graph the results of the 30 sets. He left these same instructions for another of his classes, Class B.



When he came back the next day, he saw these two graphs, showing the results of 30 sets of 50 spins for Class A and Class B:





(a) How would you describe the shape of the graph for Class A?

(b) How would you describe the shape of the graph for Class B?

(c) Mr. Shaw suspects that one of the two classes did not really do 30 sets of 50 spins, but instead just made up the data. Based on the results shown in the two graphs, do you think one class or the other is likelier to have simply made up the data?

Explain why you think this.

APPENDIX C

Class Interventions

Class Intervention #1: Data & Graphs

In the fifth week of the quarter, Steve left the topic of geometry and entered the topics of statistics and probability. Matt and I also began attending class, and we set up our audiovisual equipment on a table in the back of the room, leaving the class with seven table groups' worth of students instead of the previous maximum of eight. Steve's section had an enrollment of 30, so after Matt and I began attending the groups at the remaining seven tables had four or five students in each group. I was in attendance for eight days (spanning weeks 5 through 8), and some of the main activities

The contexts of data and graphs, sampling, and probability shared considerable overlap, and in fact Steve began with a class discussion of each of the terms which he had written on the board: Data, statistics, and probability. I have identified for each intervention a context which I thought had a more unique focus, and the dominant context in which Steve first led the class was data and graphs. The two activities comprising the Class Intervention for the context of data and graphs were called "Four Questions" and "Body Measurements," and are discussed next. It is important to remember that the purpose in sharing the details of these class interventions is to create a picture of the opportunities given to the subjects for exploring variation. Shifts in thinking from the pre- to post-instruments may indeed be attributable to the interventions, but this research does not set out to prove a treatment-and-effect dynamic. However, the environment for learning is certainly important to document, as it does offer clues as to how conceptions may be formed and influenced.

The "Four Questions" activity, as a part of the first intervention, was chosen because for two reasons. One reason is because Steve and I had each used versions of the activity with other Math 212 classes, and were therefore experienced in how it went and what it offered. The second reason is because it offered a good opportunity to discuss both average and spread in data sets. Steve therefore started the class exploration of statistics in the fifth week by having the entire class gather data from one another in response to four questions:

How many pets do you have?
How many years have you lived in Portland (to nearest half-year) ?
How many people are in your household?
How much change (in coins) do you have today?

Different groups were in charge of graphing the collected data for an assigned question in any way they wanted. Because there were more groups than questions, some questions were duplicated by different groups. However, being a very open-ended activity, the same question ended up being graphed by different groups in different ways. The graphs were all put up on poster paper in front of the room, and comparisons were made between the different types of graphs. An example of an actual graph for the number of pets of shown below in Figure 1:

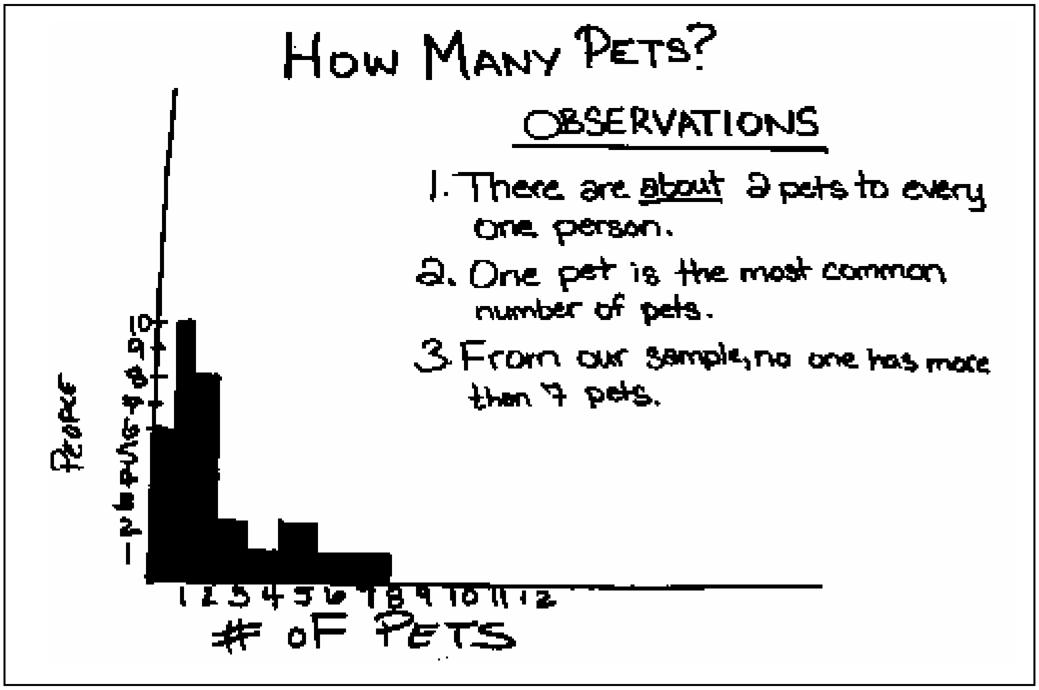


Figure 1

As an example of how, for example, the contexts of sampling and data and graphs could overlap, here is an example of a comment volunteered by one of the six cases, SP, speaking on behalf of her group about a graph showing the number of people in a household:

> SP: The most common [number] was two…No household contained more than five, and that surprised us. So we thought, "Why is that?" And then we were thinking that this class is probably not representative of the population as a whole, because of age, education level, academic status…

The idea of representativeness clearly came through from other groups also.

Having done the activity with different Math 212 classes, the variety of types of graphs the students came up with on their own was greater than I'd seen with other classes. Steve's section had come up with pie and bar charts, histograms, pictographs, and line and dot plots. A discussion ensued about the different ways that data could be presented, and also about the level of detail provided by each type of graph. One line of questioning that Steve and I prompted was the idea of what would be a "typical" value for a Math 212 student or for the class. For instance, the number of years living in

Portland ranged from less than one year to over 40 years (see Figure 2). We could find an average, but in what sense would that value be typical?

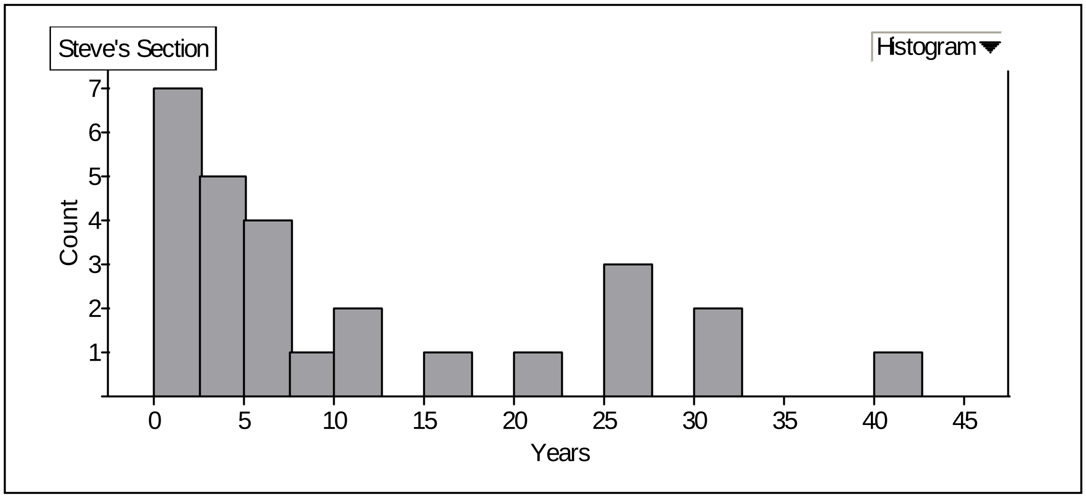


Figure 2

Figure 2 represents a computer-generated version of one kind of graph the class had in the front of the room for discussion purposes. Another table used a pie chart for displaying the data for number of years in Portland. The ideas that Steve was able to draw out or introduce included the definitions of mean, median, and mode, and how those measures appealed to a sense of average. He also highlighted the term "outlier" for the maximal value. The class picked up on how a measure like the mean doesn't even have to be an actual data point, as in Figure 2, where the mean is 11 years. The class expressed some dissatisfaction in claiming, for instance, that the typical student from Steve's section had lived in Portland for 11 years, when (a) No one actually reported the value of 11 years, and (b) More responses seemed to be under 11 years as opposed to over 11 years (the median in Figure 2 is in fact 5.5 years).

The main point is that the tension between centers and spread of data emerged as theme of discussion even from the very first day of doing statistics. Steve then went on in the next class session to investigate a model for obtaining the mean by balancing out stacks of tiles, and he also worked with the class to create and use boxplots. Particular attention was paid in the class discussions to what kind of information boxplots showed and what kind of detail they obscured.

The second activity in the class intervention focusing on the context of data and graphs was "Body Measurements", which was also selected for the reason that it was a well-rehearsed activity for Steve and me. More importantly, a similar activity was to be used for the NSF-sponsored project (mentioned earlier) that looked at conceptions of variation in middle and high school classes. Having been a part of that project, Matt and I had agreed that the activity was useful for prompting thinking about variation. In particular, causes of variation in a repeated-measurement scenario get explored in Body Measurements, as do issues of confidence in situations involving measurement.

The introduction to Body Measurements came towards the end of the first class session of week 6, with about 30 minutes left in class. By way of introduction, we discussed scenarios that we'd already seen in class whereby the data varied. For example, the data in response to the four questions shared earlier had varied. We also listed things we thought would not vary, such as the time class started each day, or the amount it costs to park on the street for an hour during classtime. I offered the idea that some of our body measurements – such as armspan or head circumference – would not vary during a short interval of time. Thus, the task was given to gather several sets of data: First, everyone in class was going to measure Matt's armspan. Second, everyone had a "personal data sheet" which required their own armspan, height, handspan, head circumference, and pulse rate per minute. The only directions given were that pulse had to be counted out for a full minute, not some shorter interval and then multiplied by an appropriate factor.

What happened next mirrored in many ways what Matt and I saw later in the middle and high school classes. Armed with meter sticks and tape measures, the Math 212 students did find partners to help gather their own personal data, and in many cases the measurements were carried out very casually. For example, head circumferences were measured above the ears for some people and around the ears for others. Handspans included a natural span (meaning the subject just opened a hand as far as it would naturally go) or a forced span (meaning that the subject found some way to open their hand even farther, such as pushing the open hand against a table). What was really interesting was the measuring of Matt's armspan. Since none of the measuring tapes or meter sticks would singularly cover the armspan, subjects were forced to find a way of compensating.

The typical way of measuring was to start at one side and measure across Matt's back until, for instance, the measuring tape ended. Then subjects would hold a finger at the ending spot, or even affix their gaze to the ending spot, and then start a new measurement from that spot to the other side. For the twenty-seven students that we had in attendance that day, the time needed to gather the measurements was sufficiently lengthy to make Matt's outstretched arms sag after awhile. Eventually he would just put up one arm at a time as needed. There were so many sources of error in just the 15 minutes of data gathering that it was not possible to list them all, but it is interesting that many of the same issues arose later when Matt and I watched the activity done in the middle and high schools. I collected all the personal data sheets, which also had each subject's measurement of Matt's armspan, and then class ended for the day.

Because the students had already practiced making different kinds of graphs in class, I decided to type up the class results and distribute sheets of the data and some graphs during the next class session. The main part of the class discussion was on interpreting the graphs. For example, the graph for Matt's Armspan is shown in Figure 3 below. Questions that we asked of the students included: Why are the measurements different? What can you conclude about Matt's armspan? How confident are you about Matt's true armspan? Several comments emerged to show that students knew many causes of variation in the repeated-measurements situation. As to conclusions about Matt's true armspan, there were different ideas expressed with different degrees of confidence. For example, the whole class seemed very confident that the true measurement was captured within the range, and some members felt that the mean (about 76.5 cm) was the true value. Other students liked the mode (which was also the median value of 76 cm) because they felt that most students must have done the measurement correctly.

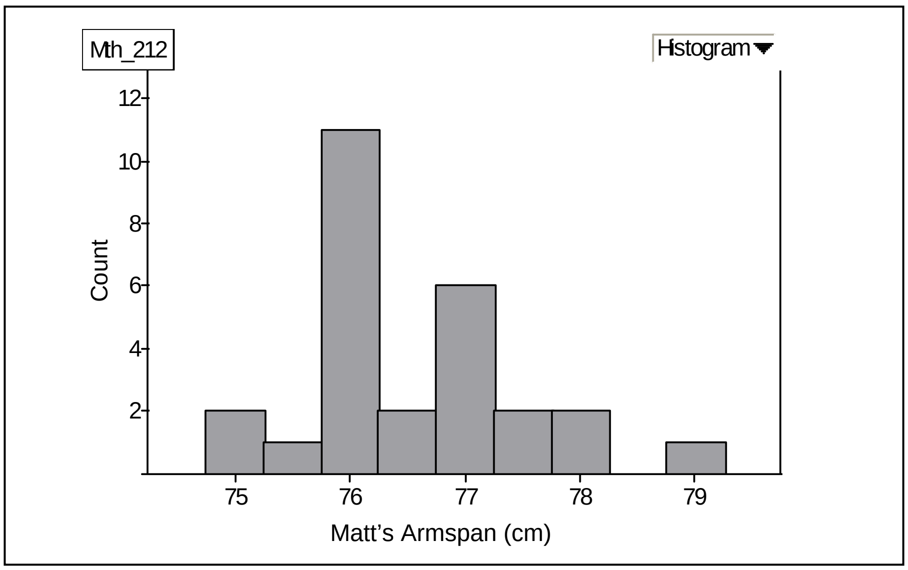


Figure 3

A big idea to come out of the discussion of Matt's armspan was that it was harder to identify a single value as being true or "accurate", but it was easier to talk about intervals of values where the true value might lie. Then in talking about the armspans for the whole class, again there was a similar discussion as we had for the "Four Questions" about what was a typical measurement for the class. Only a couple of students made a connection between the variation shown for a single measurement (Matt's armspan) and what that variation implied about the accuracy of the classwide measurements. The comment that came out in class were to the effect that, since Matt's armspan was "fuzzy" (meaning hard to pin down with precision based on the data), so too should each of the classwide measurements be "fuzzy" as well.

Class Intervention #2: Sampling

In the seventh week of class, we spent most of both class sessions doing two activities that comprised an intervention focused on the context of sampling. The two activities, “Known Mixture” and “Unknown Mixture,” were selected because Matt and I had participated in the activities at six other schools as a part of the NSF-sponsored project mentioned previously. We had seen how effective the activities were in drawing attention to variation, and decided to follow the same essential plan for carrying out the activities as had been enacted in the middle and high schools. The sampling activities had evolved in the NSF project through meetings with classroom teachers and university researchers with an interest in promoting thinking about variation. Steve had done similar activities in previous Math 212 classes.

For the Known Mixture, we started with a general discussion of what were samples, who uses samples, and what samples were good for. Then the following scenario was given as a part of a handout (see Figure 4):

<u>Scenario for Known Mixture Activity</u>

The band at Johnson Middle School has 100 members, 70 females and 30 males. To plan this year’s field trip, the band wants to put together a committee of 10 band members. To be fair, they decide to choose the committee members by putting the names of all the band members in a hat and then they randomly draw out 10 names

Figure 4

Individually, the Math 212 students considered how many females they might get in a single draw of ten names (referred to a one sample). Then they wrote down what they thought they might get in six samples, and finally they made predictions for thirty samples. It was made clear that multiple samples were done with replacement. After making the above predictions individually, the students talked within their groups, coming to a group consensus about what they expected for 30 samples, and then they brought their numbers up to the overhead. I wrote all seven groups’ numbers down, and they are shown in Table 1 on the next page.

In discussing similarities and differences in what the groups had predicted, students noted that all groups had highest numbers of females between 6 and 8, and the mode for almost all groups was at 7 (which corresponded to the proportion 7/10 = 70/100). Also, all groups had at least some results at 10. Three groups had no results lower than 3 females in a sample of size 10, while three groups predicted at least one sample having 1 female. After discussing the predictions, the directions for actually drawing 30 samples of size 10 were given.

Predicted Results for 30 Samples of Size 10											
Number of Females	0 F	1 F	2 F	3 F	4 F	5 F	6 F	7 F	8 F	9 F	10 F
Group #1	0	0	0	1	1	1	5	10	5	5	2
Group #2	0	0	0	1	2	3	6	8	6	3	1
Group #3	0	0	1	1	3	2	5	7	6	3	2
Group #4	0	1	1	2	3	4	5	6	5	2	1
Group #5	0	1	1	2	3	3	5	6	5	3	1
Group #6	0	0	0	1	2	4	6	6	5	3	3
Group #7	0	1	1	1	2	2	5	10	5	2	1

Table 1

Plastic jars with 100 chips (30 green and 70 yellow) had been prepared, with the yellow chips corresponding to the females in the band committee scenario. Students were instructed, in their groups, to draw 30 handfuls (each of size 10) , recording the numbers of yellows in each handful before returning the chips to the jar for remixing. The subsequent sampling showed many of the same lackadaisical mixing techniques as had been observed in the middle and high schools. For example, some students would return their chips to the jar and then just give a weak side-to-side shake. Especially because the chips were flat and smooth, a sideways motion is not an optimum strategy for mixing. Other students would use the stirring technique, putting their hand in for a brief stir before drawing out their new sample. Another method was the up-and-down shake of the jar, which generally had to be tempered by the fact that the jars had no tops. Too vigorous of a shaking sometimes resulted in chips getting ejected from the jars.

Eventually, twin posters went up for each group: The top posted held the graph of that groups' prediction, and the bottom poster held the actual results. We were able to get all seven pairs of posters up on the boards in the front of the room. Figure 5 shows two of the posters.

In discussing the posters, the initial questions were "What do you notice?" and "What was surprising?". Some of the comments that followed had to do with the centers, spreads, and shapes of the graphs. For example, in comparing from predicted to actual, while almost all of the modes for the predicted graphs were at 7, only three of the actual graphs had modes at 7 (modes for the groups' actual graphs were 6, 6, 7, 7, 7, 8, 8). All of the actual data is shown below in Table 2.

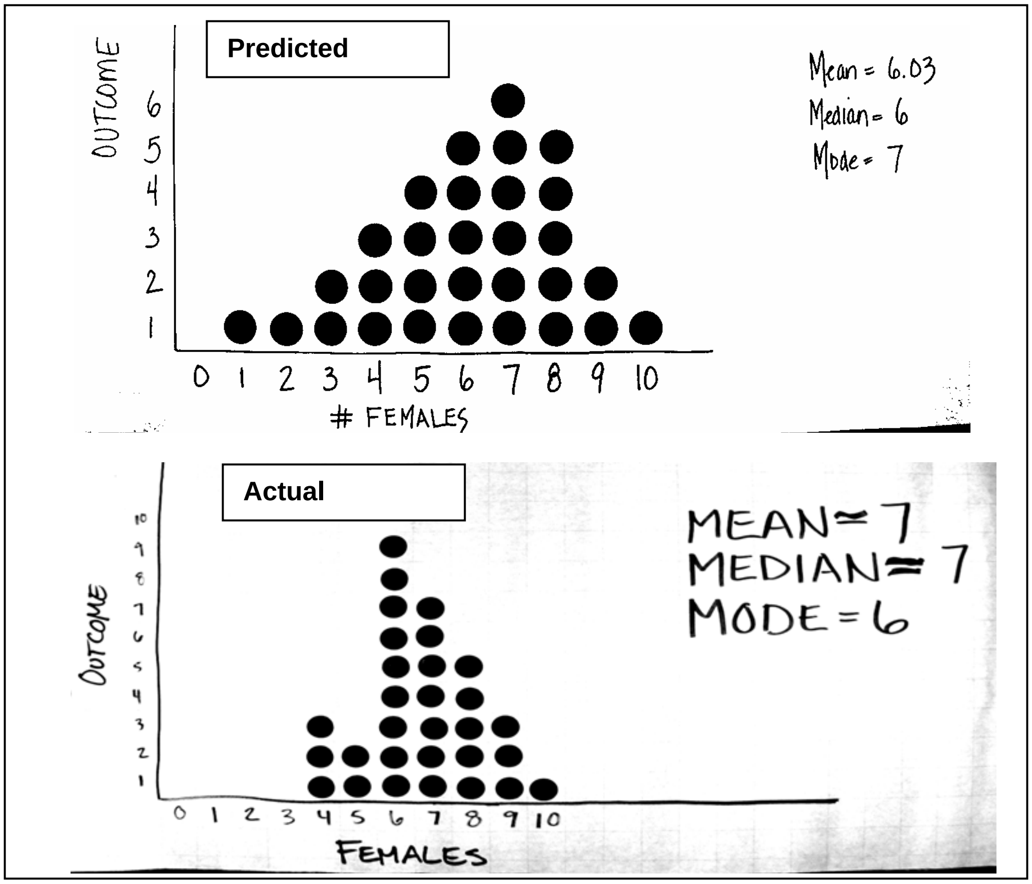


Figure 5

Students brought up how the predicted graphs were more spread out than the actual graphs, and also how the shapes of the predicted graphs all looked somewhat similar, but the actual graphs had shapes that were noticeably different from one another.

Actual Results for 30 Samples of Size 10											
Number of Yellows	0 Y	1 Y	2 Y	3 Y	4 Y	5 Y	6 Y	7 Y	8 Y	9 Y	10 Y
Group #1	0	0	0	0	1	3	8	5	6	7	0
Group #2	0	0	0	1	3	3	6	6	7	3	1
Group #3	0	0	0	0	2	4	4	7	6	6	2
Group #4	0	0	0	0	3	2	9	7	5	3	1
Group #5	0	0	0	0	0	4	7	9	6	2	2
Group #6	0	0	0	1	1	5	8	10	3	1	1
Group #7	2	0	0	0	0	5	4	6	8	4	1

Table 2

On the issue of extreme values, it was mentioned how not all groups got a 10, and how most groups did not get below a 4. Group #7, who had reported getting two handfuls containing zero yellows, did raise suspicions among other class members. They asked the members of Group #7 if they had adhered to the procedures for drawing handfuls, , and it turned out that one of my cases, GP, admitted to using his sense of touch to discern the colors of the pieces. One of GP's group members, MG, said: "If you really compared the feeling of the two [types of chips], the yellow had a sharp edge and the green had a blunt edge." GP was also shown in the interviews to have an acute awareness of the role of his hand in creating (or impeding) randomness in these kinds of sampling situations. No one else in class seemed to have been aware of the tactile differences in the chips that GP's group had exploited to deliberately draw two handfuls of zero yellow in a row. Another of my cases, in a different group, asked GP if those two non-random samples had been calculated into the mean, showing an appreciation for the effects of outliers.

One big idea that came from the discussion was that the actual posters were not all the same, and in fact looked even more different from each other than had the predicted posters. Another big idea was that just as the predicted posters had most of the data around 6, 7, and 8, so too did the actual posters. Finally, disregarding GP's results of 0, a final big idea was that the lower extremes did not seem likely to result from actual results of 30 samples. At the end of the class session, it was suggested that we could get a better "actual" graph if we combined the results from our groups.

With these ideas in mind, the next class session began with an extended time of interacting with the ProbSim software, which some students later referred to as Fathom because that was the program I had used at an earlier time in the quarter. Matt led the class through much of the investigation as he had done with the NSF project. However, we spent more time using ProbSim with the Math 212 class than we had usually done in the middle and high schools.

The first thing we did was to run many trials of 30 samples. Since the students had done and seen the results for seven trials of 30 samples, the results (which ProbSim displays quite rapidly) made sense to the class: They were essentially imagining that hundreds of groups had done trials of 30 samples, not just seven groups as they themselves had done during the previous class session. Since we had mentioned aggregating our seven groups' worth of data, Matt shifted to trials of 210 samples (corresponding to seven groups at 30 samples per group). Two things that students were quick to pick up on was how the shapes started to look very similar from trial to trial (of 210 samples in each trial). In particular, the mode had stabilized at seven, and the graphs all looked like skewed bells. However, the far extreme of 1 yellow did not always result, and 0 never appeared. Matt then did repeated trials of 500 samples, and also 5000 samples, but 0 never appeared. Students felt sure that since 0 was a possibility, it should result given enough samples. Matt, Steve, and I pointed out that collectively we had seen tens of thousands of samples of size 10, and none of them had resulted in 0 yellow. There was a considerable anticipation in the class, waiting for a result of 0, but we ended the computer simulation before that result occurred.

We then made a transition into the second activity in this intervention, which was the Unknown Mixture. It was made clear that even though we had known what was in the earlier jars, samples still had varied. Now we had larger jars, each containing 1000 chips of yellow and green, and the mixtures were known to be the same in each jar. However, the exact mixture was not known to the class (it was actually 550 yellows and 450 greens). The students were, in their groups, to decide what sample size they wanted to use (we imposed an upped limit of size twenty for all groups) and how many samples they wanted to draw. Then they were to carry out their plans, do the sampling, graph the results, and make some conjecture about the true mixture in the jar.

As an example of one groups' results and reasoning, the poster in Figure 6 shows how the mean and the median of 5 yellows helped this group decide on a prediction of 50% yellow in the jar. The groups who authored the poster in Figure 6 used a sample size of 10, and they drew 70 samples. The group who made the poster shown in Figure 7 on the next page also used samples of size ten, and they drew 40 samples. However, they obtained their "guesstimate" of 57% by looking to their mean of 5.47 yellow and also their mode and median of 6 yellow, and finding some a value that they felt was somewhat close to both 5.47 and 6, namely 5.7. Then they used the ratio of sample size to population and determined that 570 yellow chips out of 1000 total chips would correspond to 57%. They added the "margin of error" because they knew that plus-or-minus three percentage points would cover their mean, median, and mode.

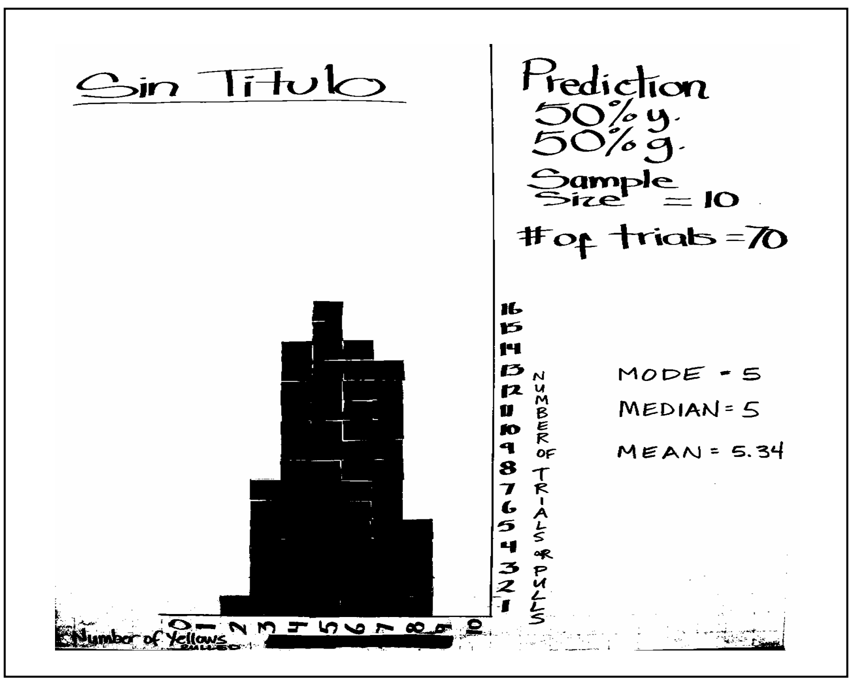


Figure 6

Other groups selected sample sizes that went as low as 6 and as high as 200, and after discussing the different estimates we decided to try for a class consensus. I introduced the idea of confidence by relating the idea to what was shown in the posters. For example. the group for the poster in Figure 5 had an estimate of 50% yellow, or 500 yellow chips, but they were not too confident the jars really did have exactly 500 yellows. The group for the poster in Figure 6 were not confident of the point estimate of exactly 570 yellow, but were fairly confident that the true value was somewhere in the interval from 540 to 600 yellow.

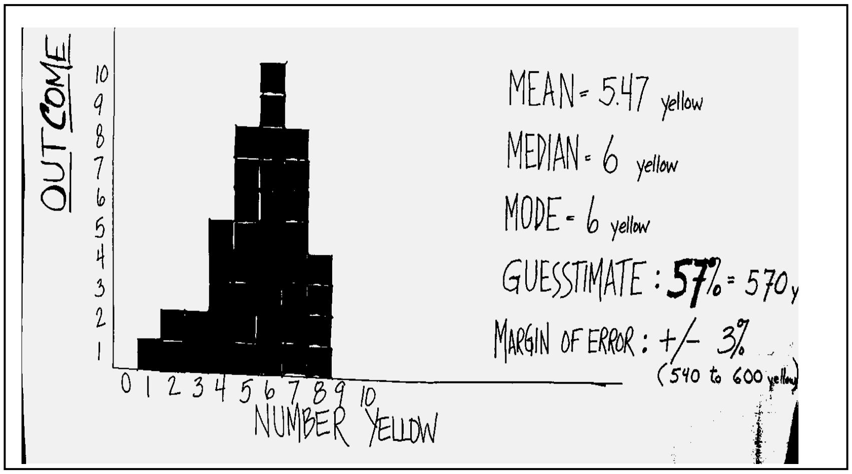


Figure 7

Comparisons across all the posters showed that the sample sizes for ranged from 6 to 20, and the numbers of trials ranged from 20 to 70. Predictions ranged from 500 to 600 yellows, with a couple of groups offering an interval. In discussion, I asked the class as a whole what would or would not be surprising to them: For example, the class expressed no surprise if the true value was 520 yellow, or 580 yellow. The class eventually came to a consensus on an interval as small as 540 yellow to 570 yellow. The big idea here was that an unknown mixture (or any other realistic sampling situation) does not mean that nothing can be said with any confidence about the mixture. In other words, the class was overwhelmingly confident that there were more yellows than greens, but not as high a ratio as 650 to 350, for example. Despite attempts to have the students accept that in real life sampling investigations, having some degree of confidence in an interval is the best that can be hoped for or expected, students still wanted to know the exact percentage, which was revealed at the end.

Class Intervention#3: Probability

There were two activities that made up this intervention, “Cereal Boxes” and “The River Crossing Game,” and these were chosen specifically because of the probability aspects involved in the activities. Namely, Cereal Boxes relies on the use of spinners and River Crossing on the use of dice as random generators, and these two activities were the main ones done in Math 212 involving spinners or dice.

Cereal Boxes actually took place in the first class session of week 2, just before we gathered data for Body Measurements. As explained earlier, there was considerable overlap in the contexts, and Cereal Boxes is a good example of this overlap. Cereal Boxes is sample-until scenario, supposing that any of five different stickers can be obtained with each box of cereal opened, and that the five stickers have equal chances of being obtained. The question is, about how many boxes would be opened to obtain all five stickers, and the situation can be simulated by using a five-region spinner. Cereal Boxes brings together probability, sampling, and data and graphs in way that also highlights variation, and that is why I chose to use the activity.

After posing the problem, guesses were taken from the class about how many spins they expected to have to do to hit every region at least once. An upper limit of 1000 was suggested, and one student suggested the upper limit was infinite. Also, a students suggested an expectation of 5 spins, the lower limit. Although this latter student was labeled an optimist by classmates, it focused attention on the probability aspects of this activity: After all, if we say there is a 1-in-5 chance of getting a region on the spinner, then after 5 spins there might be an expectation that all regions have been hit. As for the suggestion of infinite spins, it also focused attention on the idea that there was a chance of never hitting a certain region.

After discussing expectations, working individually at their tables, each student performed 10 trials (one trial was defined as finding the least number of spins it took to hit all five of the regions at least once). SW, for example, whose initial guess was 105 spins, obtained the following results for 10 trials: 9, 5, 20, 6, 14, 17, 6, 14, 10, 9. I asked her if, after looking at the results of her 10 trials, 105 spins was still about what she would expect:

> SW: I think I might lower it a little bit, maybe to 70, in the 70s. But I think it would take quite a bit, it might even be higher than that [105]. Because, I mean, you’re not just dealing with numbers, then, you’re dealing with, you know, they want you to buy as many boxes of cereal as possible, so they’re going to spread it out a little bit more.

The class did raise the issue of what a cereal company's actual distribution would be like, and Steve re-focused the attention on the assumptions of the simulation. Thinking strictly in terms of numbers of spins to get the different regions on the five-spinner, new expectations were made in light of the trials gathered. Instead of ranging from 5 to 1000 spins, the second round of predictions for an expectation ranged from 7 to 25.

Another key issue arose when Steve was asking the class to notice how the second round of predictions was tighter than the first round, and he asked "Which answer is right?" After someone said that there was no right answer, Steve steered the conversation to the idea that some predictions seemed more reasonable than others, and a part of the discussion included the idea of aggregating results to get even better predictions. After aggregating individual results at their tables, there were seven group predictions based on 40 to 50 trials, and the predictions ranged from 7 to 11 spins. Students did want to aggregate the whole class results sensing that more trials offered an increasing better idea of what to expect. Instead of taking the time to graph all the data, I brought out the Fathom software. Figure 8 shows the results for 150 trials:

Steve talked about where the upper and lower 10% of the data was, and also used the Fathom graphs to talk about boxplots and distribution of data. While the mode in the data set for Figure 8 is 9 and the median is 10, the mean is 11.7 and quite close to the expected value in this situation. But what students were aware of from having done their own trials was the variation in this situation, and how the chances of getting a certain region on that spinner were not guaranteed.

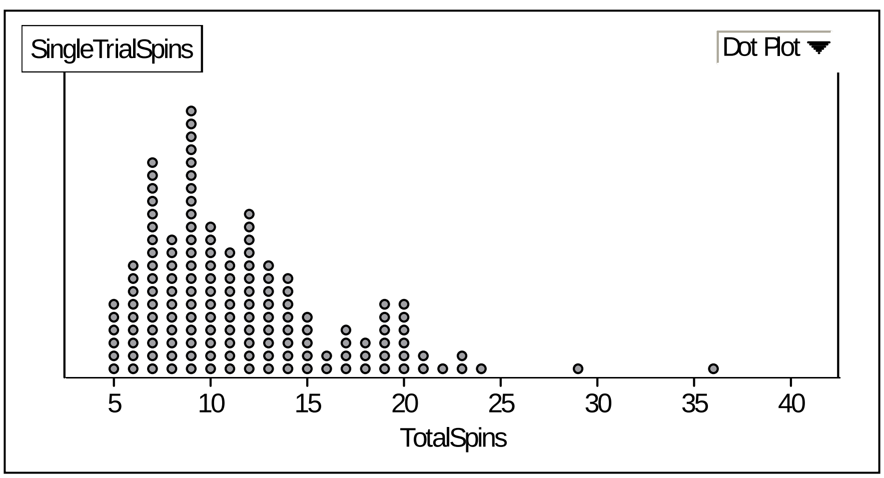


Figure 8

In Figure 8, students knew that the meaning of the upper extreme was that one region on the spinner got avoided for 34 spins, only receiving a hit on the 35th spin. Cereal boxes, although modeling a sample-until scenario, did afford students good opportunities to focus on probabilities involving the spinners, which was among the reasons I had wanted to include it as a part of the intervention for the context of probability.

The second activity for this intervention, the River Crossing Game, involved finding the sum of two dice. Credit for this activity goes to the *Math and the Mind's Eye* curriculum (Shaughnessy & Arcidiacono, 1993). Using two players, each player got 12 chips to place on their side of a "river", along spaces marked 1 through 12.

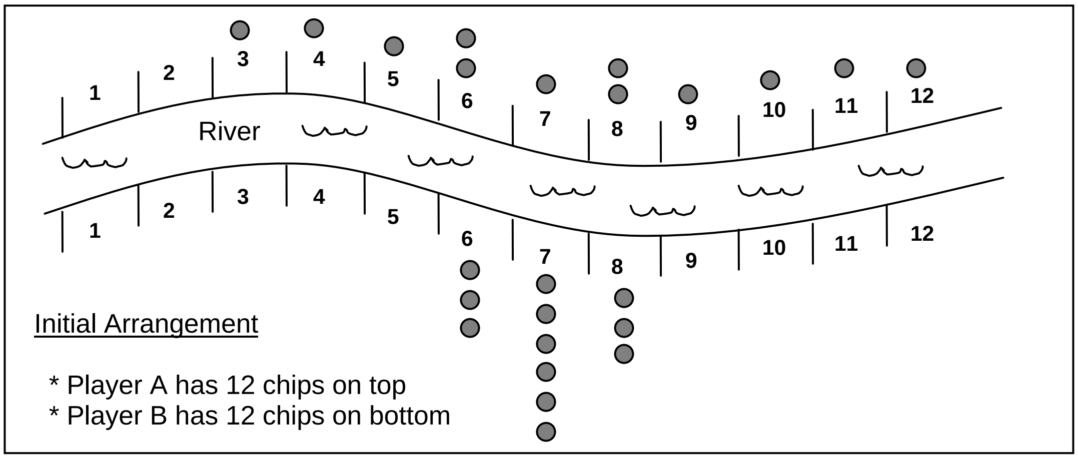


Figure 9

After configuring their chips in an initial arrangement (see Figure 9 for an example of two players' initial arrangements), players took turns tossing a pair of dice. If either player had any chips on the space showing the total for the dice, one chip could "cross the river" and be removed from the board. The winning player was the first one to remove all the chips on his or her side. For instance, in Figure 9, if the dice resulted in a sum of 10, Player A on top could remove one chip. If the dice showed 8, Player A and B could each remove one chip.

Teams of students played several games, and kept track of the results of each toss of the dice on a dot plot: Sums were given along the horizontal axis, and students could just make a mark showing the sum obtained. Some students' initial arrangements changed from game to game. For example, seeing from their dotplots how sums of 6, 7 and 8 tended to occur more than other sums, some students put most of their chips on those spaces. Other students continued to spread out their chips, feeling that in the course of tossing the dice, they would get a sum of 2 or 12 for example.

After a few games, the dotplots from the teams were put up on the board in front of the class, and each graph showed well over one hundred tosses of the dice. Steve led the class in a discussion about the graphs, their shapes, and what the class as a whole might expect. Despite the variation shown in each graph, students volunteered that what they felt was most likely to happen for a sum would be 6, 7, or 8. Some students kept volunteering 7, but Steve made a distinction between the knowledge that some already held (that 7 was most likely) and what the class' experimental data suggested. Most of the class thought that 8 came through as being most likely, based on class results. Steve pointed out that "there are trends, but there's still a lot of variation from group to group." Looking across the graphs, the idea of a symmetrical theoretic distribution also emerged, and the class later went on to look at the theoretical probability for the sum of two dice.

After the idea about 7 being the theoretically most likely sum emerged, Steve also discussed what the class thought might be an optimum initial arrangement for playing the River Crossing Game. An example of two strategies, informed by having now had experience playing and by looking at the theoretical distribution, are given in Figure 10 below:

```
                x  x  x                                      x  x
                x  x  x                                   x  x  x
             x  x  x  x  x  x                       x  x  x  x  x  x  x
 1  2  3  4  5  6  7  8  9  10 11 12       1  2  3  4  5  6  7  8  9  10 11 12
```

Figure 10

A question arose as to why people don't just put all their chips on 7, since that sum is most likely. Steve brought out the connection that putting all the chips of 7 was making a statement of expectation of no variation. He pointed out how the dotplots of results showed not only what happened most, but also gave an idea of the kind of variation we could expect as well.