Contents

Chapter 1

Probability

"It is a truth very certain that when it is not in our power to determine what is true we ought to follow what is most probable." - Rene Descartes, mathematician and philosopher.

"Misunderstanding of probability may be the greatest of all impediments to scientific literacy." - Stephen Jay Gould, biologist.

1.1 Introduction to Probability

Your group should now have two coins. Consider a game where there are two teams, one called the Brewers and one called the Cubs. You flip the coins, and if the faces that come up match (i.e. both heads or both tails), the Brewers get a point. If the faces are different, the Cubs get a point.

1. Suppose you were flip the coins ten times. Which team do you think will have the most points? Explain why. Make an exact prediction (like 10-0, or 7-3). For what reasons did you come up with that prediction?

2. Suppose you were to flip the coins until one team had 10 points. Which team do you think would win? Explain why. Is the reason the same as the reason to the previous part? Make an exact prediction for the score. What reasons led you to choose that prediction?

3. Now flip the coins ten times. Did the outcome match with your prediction? If so, was it a coincidence, or do you think your prediction was a good one? If not, why not? Was your prediction a bad one, or was it close?

4. Do the results of that experiment make you change your prediction in part 2? If so, what made you change your prediction? What is your prediction now?

5. What do you think the "probability" is of the Brewers getting a point? Of the Cubs? Think carefully about how your got this answer. What reasoning/information was it based on? What is your definition of probability that gave you this answer?

1.2 Modeling Problems

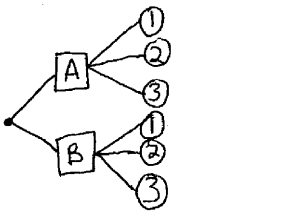
1. (a) If a basketball player with a 58% free throw average takes two free throws, what are the chances they will make both of them? Be prepared to explain your reasoning to the class.

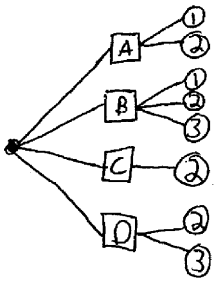
   (b) Now suppose that if they make the first shot, their average for the second shot goes up to 80%. What is the likelihood they will make both shots in this scenario?

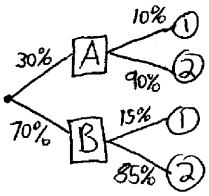
   (c) If you haven't done so already, represent the data of the previous question with a tree model, and with an area model.

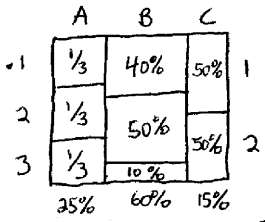
2. There are 6 possible pairs of dice out of A,B,C,D. When you get to this problem, tell me, and I'll give you 2 of the possibilities. Determine which die you think will win in each case, and how much of the time.

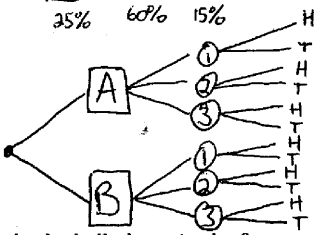
3. If given a tree model, determine the area model that represents the same data. If given an area model, write down the corresponding tree model.











4. For the basketball player in the first part of the first question - 58% average, regardless of what shot it is - consider a one-and-one scenario. That is, they get to take a second shot, but only if they make the first one. In 100 trips to the free-throw line, how many points would you expect them to score total? Give a specific model!

1.3 Why Multiply?

1. Suppose you have an urn with 15 red balls and 5 white balls.

   (a) What is the probability of pulling a red ball? White ball? Call these Pr(Red) and Pr(White), respectively.

   (b) Suppose you are going to pull a ball, replace it, then pull another ball. What is the probability of pulling a red ball then a white ball? Call this Pr(Red & White) Create both tree and area models for this experiment. Can you represent this in terms of the probabilities in the previous part? That is, can we calculate Pr(Red & White) from Pr(Red) and Pr(White)?

   (c) Suppose you are going to pull a ball, then pull another ball without replacing the first ball. What is the probability of pulling a red ball then a white ball? Again, call this Pr(Red & White). Create both tree and area models for this experiment. Can we calculate Pr(Red & White) from Pr(Red) and Pr(White) as in part b)?

   (d) Can you explain why in part b) you could multiply the individual probabilities, but you couldn't in part c)? Try to be as specific as possible. How did the change from the experiment in part b) to the experiment in part c) change the problem?

2. Suppose your urn has 10 balls, each with a color and number. Balls 1-5 are red, and 6-10 are white.

   (a) What is the probability of pulling a red ball? White ball? Even numbered ball? Odd numbered ball? A ball with number of 8 or higher? Call these Pr(Red), Pr(White), Pr(Even), Pr(Odd), and Pr(≥ 8).

   (b) What is the probability of pulling a ball that is red or has a number 8 or higher? Call this Pr(Red or ≥ 8). Can you express this probability in terms of the probabilities you found in the part a)? That is, can we calculate Pr(Red or ≥ 8) from Pr(Red) and Pr(≥ 8)?

   (c) What is the probability of pulling a ball that is red or has an odd number? Call this Pr(Red or Odd). Can you express this from Pr(Red) and Pr(Odd) like you did in part b)?

   (d) Can you explain why in part b) you could add the individual probabilities, but you couldn't in part c)? Try to be as specific as possible.

1.4 Conditional Probability and Independence

For events E_1 and E_2, recall that the probability of E_1 *given* E_2 is called the conditional probabilty of E_1. This is written as $Pr(E_1|E_2)$. We can use this to help us understand when two events are independent by the following rule: Events E_1 and E_2 are independent if $Pr(E_1|E_2) = Pr(E_1)$. That is, knowing that event E_2 has occurred doesn't make event E_1 any more likely or less likely. Use this notion to determine whether the following events are independent.

1. E_1 : a selected playing card is an ace; E_2 : a selected playing card is a club.
2. E_1 : a selected playing card is a club; E_2 : a selected playing card is a black card.
3. E_1 : it is sunny out tomorrow; E_2 : it is raining tomorrow.
4. E_1 : the Badgers win their next football on Saturday; E_2 : the starting running back runs for over 150 yards and two touchdowns.
5. E_1 : a randomly selected Fortune 500 CEO has blue eyes; E_2: that person's secretary has blue eyes.

1.5 Expected Value

Consider the following two scratch ticket games.

Easy Money

Payoff	Odds
\$2	1 in 3
\$5	1 in 4
\$10	1 in 5
\$50	1 in 10

Big Bucks

Payoff	Odds
\$10	1 in 8
\$20	1 in 10
\$50	1 in 20
\$100	1 in 50

1. Which game would you rather play? Why?
2. Which game is a better investment? That is, if you could play each game many, many times, which game would win you the most money?
3. Determine how much the lottery commission would have to charge for each ticket in order to break even in the long run.

In number 3, the value you found for each game is called the **expected value** of that game. You can find the expected value for any probability experiment that has numerical outcomes (here it was amount of money won - it's a little trickier to find expected value when the outcomes are apple versus grape Blowpops). Roughly, the expected value tells you the average value of the outcomes if you were to repeat the experiment many times.

4. On a certain slot machine has 3 windows. In each window, you can get either a cherry or a lemon. In each window, a lemon is twice as likely to come up as a cherry. If you pull and get all cherries, you win \$10. If you pull and get all lemons, you win the booby prize of \$1. Determine the amount of money you should charge per pull so that this machine breaks even in the long run.
5. While working on the scratch ticket problem above, Jerry announces that the expected value for the Easy Money game is \$2. When asked why, he says, "you'd *expect* to win that prize the most often. If you played a whole bunch of times, you'd get \$2 more often than any other prize, I mean, *on average* you'd get \$2." How would you respond to Jerry's reasoning?

Chapter 2

Algebra

"We may always depend on it that algebra, which cannot be translated into good English and sound common sense, is bad algebra." - William Kingdon Clifford, mathematician and philosopher

2.1 Equality

The idea of equality has been around for centuries, but it is a very tricky concept for some elementary school students. According to Liebniz's law, two entities are the same exactly when everything you can say about one, you can say about the other. However, as we shall see, not all uses of the equal sign "=" are the same (that is, not all equalities are equal). Below is a list of various equalities. Think about them carefully. How is the use of = used in each case? How are the uses of = different? How are they similar? Some of the cases look like they might not be different at all, but think about how you think about them. In other words, when you saw these cases for the first time, what did you immediately think about them?

1. $5x + 3 = 8$

2. $x^2 = 1$

3. $x^2 = -1$

4. $x^2 + y^2 = 1$

5. $a + b = b + a$

6. $1 + 3 = 4$

7. $1 + 3 = 5$

8. $E = mc^2$

9. $F = \frac{Gm_1m_2}{r^2}$

10. $A = \pi r^2$

11. $y = mx + b$

12. $y = ax^2 + bx + c$

13. $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$

14. $e^{i\pi} + 1 = 0$

Now that we have equality understood (hopefully), let us see how it works in problems. You will undoubtedly find these problems easy, but the point is not to find the answer, but to think about **how** you got your answer. Carefully think about each step. Think about how you are using the equal sign. Also, think about other ways you (or a potential student) might solve a problem. For example, you may use subtraction to solve a problem. What if you didn't know how to subtract two numbers (in elementary school, most student are taught addition before subtraction).

1. $3 + 4 = \_\_\_$

2. $\square = 5 + 2$

3. $7 + 5 = \square + 3$

4. $100 + \square = 102 + 3$

5. $4 = \_\_\_ + 2$

2.2 Role of Variables

1. Solve the following algebra problems, or decide whether the given statement is true or false, as appropriate.

   (a) What is x, if $x + 9 = 7$?

   (b) If $y = 3$, what is $y + 5$?

   (c) If $y = 3$, is $y + 6 = 8$?

   (d) If $y + 6 = 8$, is $y = 3$?

   (e) If $x^2 + x - 2 = 0$, is $x = 2$?

   (f) If $x = 2$, is $x^2 + x - 2 = 0$?

   (g) $22 - 5 = 15 + 2$

   (h) $x^2 + 2x + 1 = x^2 - 2x - 1$

   (i) $(x + 1)^2 = (x - 1)^2$

   (j) $x^2 + 2x + 1 = (x + 1)^2$

   (k) $x^2 - 2x - 1 = (x - 1)^2$

   (l) $y = 3x + 2$

   (m) $y + 3 = (x + 3)^2$

   (n) $y = 3x + 2$ AND
   $y = x$

(o) $y + 3 = (x + 3)^2$ AND
$y = x$

(p) $y = 3x + 2$ AND
$y = x$ AND
$y = -x + 2$

2. (a) For which questions did you have to decide if it was true or false? For which questions did you have to solve? For which questions did you simplify?

(b) Restate (1a) as a question like (1c), and vice-versa.

(c) What is the difference between an equation and an expression? What is a formula? Give an example of each.

(d) What are the differences between solving, simplifying, and evaluating? Which did you do for each question in #1? Can you identify this from the way the question is asked?

(e) Is the statement $x = 3$ true? Is it false? Do you agree that it is a statement?

(f) What about $x^2 + x - 6 = 0$?

(g) $(x + 3)(x - 2) = 0$?

(h) $(x + 3)(x - 2) = x^2 + x - 6$?

(i) $y = 4x - 3$?

(j) For which questions were there more than one answer? Were there any questions for which there was no answer?

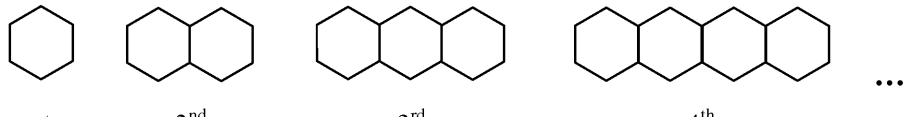
3. A middle school textbook has the following word problem: The Westons are going to build a fence around their garden, which has the following dimensions. Write an equation that describes how much fencing is needed. Then use it to solve in the case $x = 30$.

   First, answer the question. Now - what types of symbols did you use in your equation? Do you consider "ft." to be a variable? If not, how is it different from a variable?

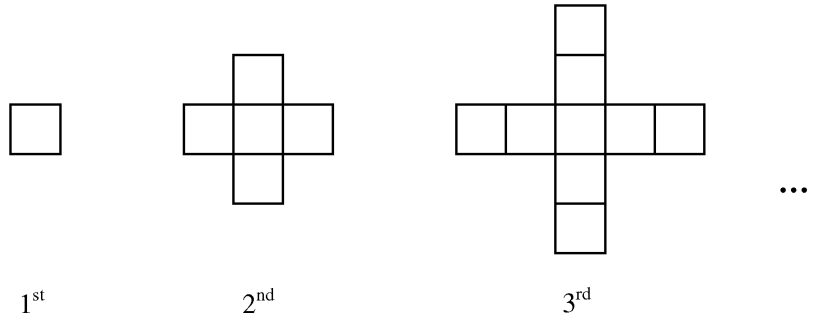
4. What do the different symbols mean in the expression y=mx+b? Are they variables? If not, what would you call them?

2.3 Patterns

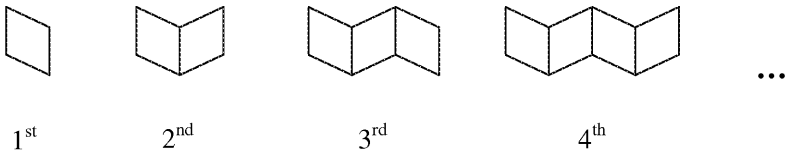
1. Write a rule using words and/or symbols *for the perimeter* (the length around the outside of the figure) in the n^{th} stage (all the sides of each hexagon are 1 unit long). Explain how you got your rule.



2. Write a rule using words and/or symbols *for the number of blocks* in the n^{th} stage. Explain how you got your rule.



3. Write a rule using words and/or symbols *for the perimeter* of the n^{th}stage (the sides of each rhombus are 1 unit long). Explain how you got your rule.



After sharing our rules as a class, for each pattern from the previous Patterns worksheet do the following:

1. Work in groups to justify two different rules *in terms of the figures*. A thorough justification should allow you to point to each part of the rule (numbers, operations, variables) and explain what they have to do with the figure.
2. Does this way of thinking lead you to see any other rules that are not already listed on the board?

Share explanations with the class and possibly consider some new rules.

Discuss and work in groups:

1. Why is connecting to the figures helpful?
   (a) For your understanding?
   (b) For your justification?
   (c) How are your answers to (a) and (b) related (i.e. what is the relationship between justifying the rules and understanding them and the problem?
2. Choose one of the patterns and create both a table and a graph. Choose one of the rules and explain how each part of the rule (as with #1 above) shows up in/is related to both the table and the graph.

1. Find numbers that fit into the pattern.

   (a) 1, 2, 4, ___, 16, ___, ...

   (b) 1, 2, 4, ___, 11, ___, ...

   (c) 1, 4, ___, 16, ___, ...

   (d) 1, 2, 3, ___, 5, 6, ...

   (e) 1, 2, 3, ___, 8, 13, ...

2. Find numbers that fit into the table.

   (a)

1	1
2	8
3	27
4	___
5	125

   (b)

3	4
5	8
___	12
9	___
11	___

   (c)

0	___
5	10
10	20
15	30
___	___

3. Imagine a very, very long hallway in a school, with lockers all the way down one wall. You are at the beginning, and you see that the lockers are numbered: 1, 2, 3, 4, 5, 6, ... The students arrive to school, and line up outside. To practice counting, they do the following: The first student enters, and counts out loud. As she says the number of each locker, she opens it, and leaves it open. Now the second student is comes in. This student has been given the even numbers. She walks down the hall as she counts out the even numbers, and as she reaches each even-numbered locker, she closes it again. When she's done, all the even-numbered lockers are closed, and the odd ones are open. The third student is assigned multiples of 3, so he counts: 3, 6, 9, ... Locker #3 was open, so he closes it, Locker #6 closed, so he opens it, he closes Locker #9, etc. The students continue, just like this. The fourth student counts multiples of 4, etc., and whenever they count an open locker, they close it, and vice versa, a closed locker gets opened.

   When all the students have passed, which lockers are left open?

2.4 Multiple Patterns

Here are the first two stages of a pattern:

Do the following parts on your own, without consulting other members of your group:

1. Draw the next two stages of the pattern, and count the number of circles in those stages.

2. In going from one stage to the next, how many circles are added? That is, how many more circles are there in stage $n + 1$ than in stage n?

3. Write out a recursive formula for the number of circles. Can you develop a closed formula? If not, how is your pattern different from the patterns where we could develop a closed formula?

Now, reconvene as a group to discuss the next parts:

4. Compare the patterns each of you drew. Did you all have the same pattern?

5. For each pattern your group came up, explain how the patten was created. That is, how did the person who made the pattern understand the first two stages of the pattern?

6. Think back to how we discussed the need to justify the rules we came up with in terms of the patterns. Does this activity reinforce that notion? As a teacher, why is it important to ask students justify a rule in terms of the pattern?

2.5 Qualitative Graphs

We have dealt with graphs often in the past: graphing lines, plotting points, etc. Here, we are going to take a more basic approach to understanding graphs.

1. Consider the following story: Dion, a math graduate student, lives on the south side of Madison. This morning, at 10, he walked through the arboretum. After returning home, he realized he was late for his math class, so he ran to the bus stop, just in time to catch the bus. After class, he walked down State Street, stopped for lunch, walked back to the math department, and biked home (he had left his bike at the department).

   (a) Draw a graph that depicts this story. **Leave space on all sides of your graph, as you will change it later**.

   (b) What are the units on your axes? Be very clear which unit goes on which axis, and why your graph explains the story.

   (c) Suppose that before Dion leaves for his walk, he goes to a coffee shop. Change your graph to depict this.

2. You are given a set of graphs and a number of situations. For each situation, choose a graph which is the best fit. Explain why. What are the units on each axis?

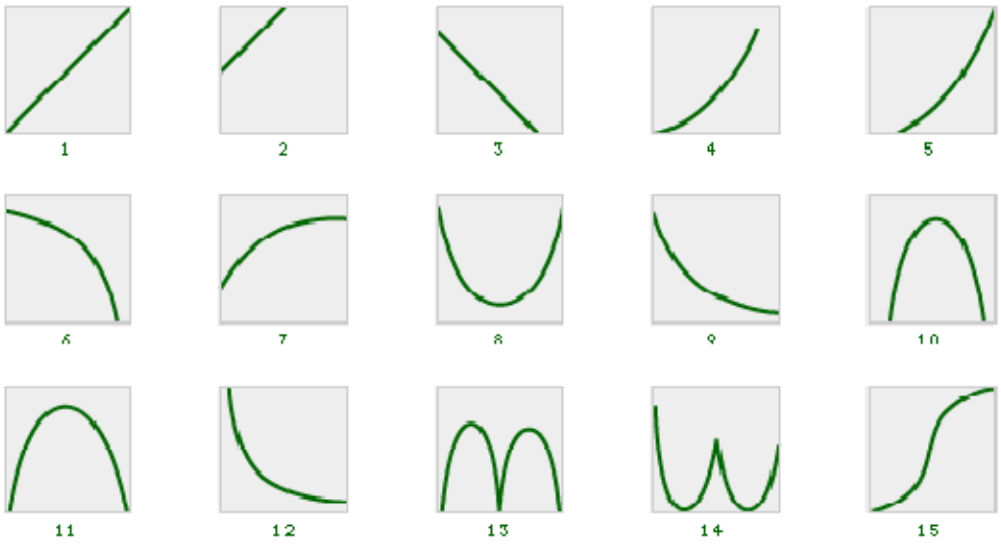
   (a) I really enjoy cold milk or hot milk, but I loathe lukewarm milk.

   (b) Prices are now rising more slowly than at any time during the last five years.

   (c) The smaller the boxes are, the more boxes we can load into the van.

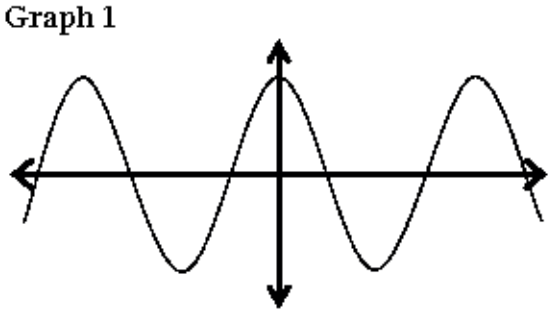
   (d) After the concert there was a stunned silence. Then one person in the audience began to clap. Gradually, those around her joined in, and soon everyone was applauding and cheering.

   (e) If the price for movie admission is too low, then the owners will lose money. On the other hand, if admission is too high, then few people will attend, and again the owners will lose. A movie theater must therefore charge a moderate price in order to stay profitable.

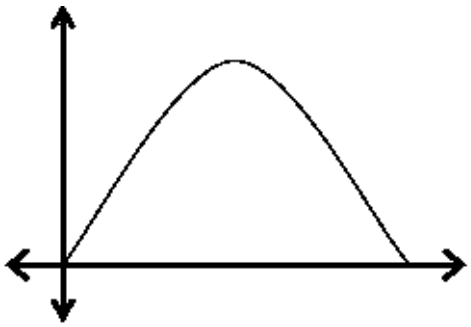


3. For the following graphs, tell a story. Be sure to say what the axes represent, and do not use a story we have already used in parts 1 and 2.

   (a) Graph 1



   (b) Graph 2



   (c) For Graph 3, choose one of the graphs you did not pick in problem 2.

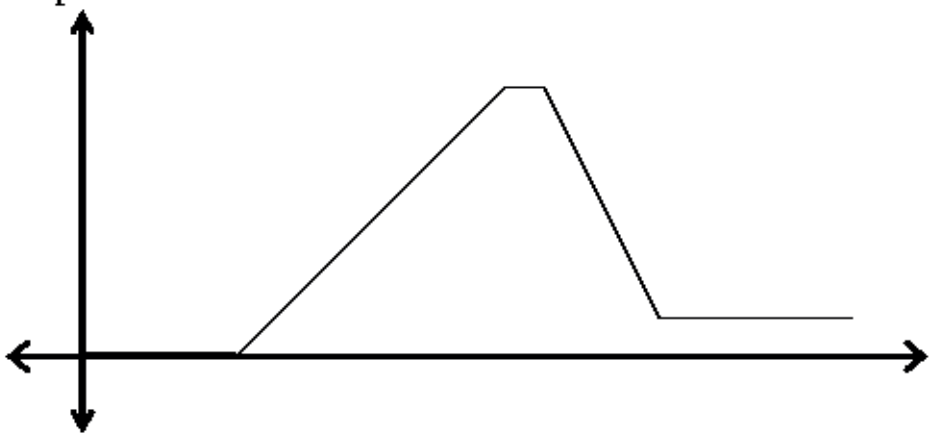
4. Consider the following story: Julia was biking for a short distance at a pretty steady pace before arriving at a hill. She biked up the hill and then sped down it and continued on a flat surface.

   (a) Each of the two graphs below can be used to describe the story. Label the axes so that the graphs make sense. Label key events in the story on the graph.

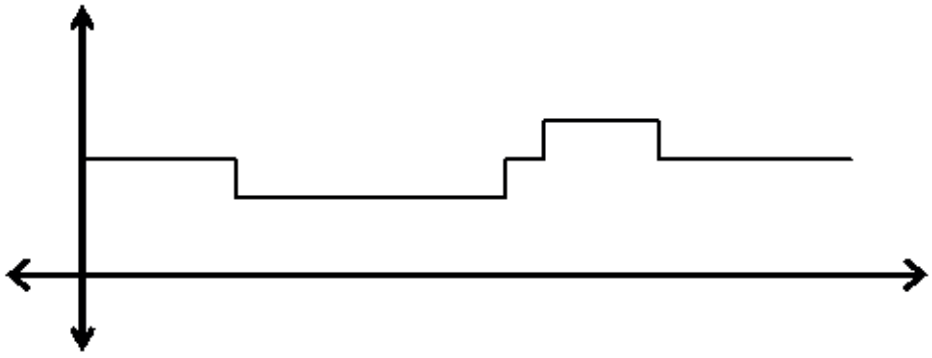
   (b) For each graph, what additional information can you get from the graph? Modify the story to fit that information.

   (c) Draw an additional graph that could go with the story.

Graph 1



Graph 2



2.6 Proportional Relationships

1. If you put 1 Tbsp of salt in 5 cups of soup, and then decide to increase the recipe to 7 cups of soup, how much salt should you add?

2. Which will be saltier, 7 tsp salt in 10 cups of broth, or 11 tsp salt in 16 cups of broth?

3. Which will result in saltier soup, adding 1 tsp salt, or reducing the broth by 1 cup? Does it matter which soup we're talking about?

4. If you are traveling 12 ft/s, how far can you go in 3 hours?

5. What is 3 ft/s in mph?

6. (a) The exchange rate today for dollars to euros is .628431. Which is worth more, \$150 or €100?

   (b) The exchange rate for dollars to British pounds is .505118. Which is worth more, €100 or £100? What is the exchange rate from euros to pounds?

7. How far away is a destination that shows as 4.5" on a 1:25,000 scale map? Would you say this scale map is appropriate for showing the downtown of Madison? The state of Wisconsin?

8. How big would the United States be on such a map? What would an appropriate scale be for a map of the whole US?

2.7 More on Proportional Reasoning

1. We now know what a proportional relationship is: x and y are proportional if $\frac{x}{y}$ is always constant. Use this to figure out if the following lists of objects are proportional or not:

   (a) For a square, the length of one and side and its perimeter;

   (b) For a $30-60-90$ triangle, the length of its shortest side and its hypotonuse;

   (c) For a circle, the diameter and the circumference;

   (d) For a circle, the radius and the area.

2. A certain shade of green paint was created by mixing 2 cans of blue paint with 7 cans of yellow paint. Using this information, solve the following problems.

   (a) Which shade of green is 'bluer': the one above, or one made from 3 cans of blue paint and 12 cans of yellow paint? Discuss this question in terms of additive and multiplicative comparisons.

   (b) If we have 5 cans of blue paint and want to recreate the same shade of green, how many cans of yellow paint will we need? Answer this question using the cross multiply algorithm.

   (c) If we instead start with 4 cans of blue paint, how many cans of yellow paint will we need in order to get the same green? Answer this question without cross multiplying. Can you use a picture to explain your solution?

   (d) For each of the following, use your technique from part (c) to determine how many cans of yellow paint you will need to create the same shade of green: if you start with 6 cans of blue paint? 10 cans of blue paint? 1 can of blue paint?

   (e) Think about this technique graphically. Draw a graph with "cans of blue paint" on the horizontal axis and "cans of yellow" paint on the vertical axis. Plot your answers for parts (c) and (d) on this graph. What is the interpretation of slope in this context?

   (f) Think about this technique algebraically. Generalize your technique so that you can find the number of cans of yellow paint required for any given number of cans of blue paint. What if we started with $\frac{1}{3}$ of a can of blue paint? 2.15 cans of blue paint? π cans of blue paint?

   (g) Connect this generalized technique to the cross multiply algorithm.

2.8 Linear Relationships

Three bikers were training for a bike race, some of their information from their training is shown in the tables below.

Biker A

Time	Mile marker
12min	61
36min	67
1hr 20min	78
1hr 32min	81

Biker B

Time	Mile marker
15min	16
45min	27
1hr 0min	30
1hr 12min	35

Biker C

Time	Mile marker
24min	34
36min	39
1hr 12min	54
1hr 48min	69

Working in a group answer the following questions and be prepared to share your thinking with the class.

1. For each biker figure out if they are biking at a constant speed (at least as far as you can tell based on the data in the tables). Provide a full explanation for how you solved this problem and why that makes sense.

2. For the two riders that were biking at a constant speed:

   (a) How fast were they biking?

   (b) What mile marker did they start at?

   (c) How long does it take them to bike one mile?

Bikers A and C are going to race and they race at the same (constant) speeds they had in training. The race course is 75 miles long. Biker C is going to start one hour after Biker A.

Working in a group answer the following questions. Be sure to show how you solved each problem and be prepared to share your thinking with the class.

1. Create a graph of this situation.

   Hint: A reasonably sized, carefully drawn graph may help in this lesson.

2. Will Biker C catch up with Biker A by the end of the race? If so when and where? Try to solve this problem in two different ways.

3. When will each person finish the race?

4. How long should Biker C wait to start if they are going to finish at the same time?

5. Develop an algebraic equation/rule for each biker that tells their distance in the race at a given time.

   (a) See if you can solve question 4 using these algebraic rules.

Talking About Slope

6. What is the slope for each biker in this problem?

   (a) What does the slope tell you in the problem situation?

7. How can you "see" the slope in the

   (a) graph?

   (b) table?

   (c) algebraic rule?

8. What is the slope *mathematically*? That is, how would you describe what the slope is in terms of variables without referring to the problem situation?

Adding Biker B

Assume that Biker B only trained for 1 hour and 12 minutes.

9. What was Biker B's average speed? How did you figure this out and why does that make sense?

10. What does finding Biker B's average speed tell you?

Assume that Biker B is also in the race with Bikers A and C and that he/she bikes at her/his average training speed the whole time. Biker B starts half an hour after Biker A.

11. Will Biker B catch up with Biker A by the end of the race? If so when and where?
12. Will Biker C catch up with Biker B by the end of the race? If so when and where?
13. When will Biker B finish the race?
14. How long should Biker B wait to start if they are all going to tie?
15. Develop an algebraic equation/rule for Biker B that tells her/his distance in the race at a given time.

Biker B's Slope

16. What is the slope for Biker B?
17. How can you "see" the slope in the
    (a) graph?
    (b) table?
    (c) algebraic rule?

From fractions to linear equations, proportional reasoning is at the heart of much elementary and middle school math. Let's look at the relationship between proportionality and linearity.

1. Suppose you know that x and y are in a linear relationship with each other. Also suppose that you know the points $(5, 12)$ and $(9, 22)$ are on the graph of this relationship.

   (a) Without finding the equation of the line, find y when x is 7. Justify your reasoning.

   (b) When $x = 3$, find y without finding the equation of the line. How about when $x = 12$?

   (c) Discuss the relationship between slope and similar triangles.

2. Suppose we have factory that bought a Widget Machine. As part of the deal, the factory got 10 widgets up front and their machine makes 9 widgets in an hour and a half. Is this a proportional situation? That is, is the ratio between 'total widgets' and 'time' constant?

3. (a) Draw a Venn diagram to demonstrate the relationship between 'proportional relationships' and 'linear relationships'.

   (b) Explain how you can tell the difference between the two in terms of equations as well as in terms of graphs.

   (c) In words, one might describe a proportional relationship by saying that the ratio between x and y remains constant. Fill in the blanks: A linear relationship is such that the ratio between _____ and _____ remains constant.

2.9 Non-Linear Relationships

Chapter 3

Data Analysis

"Aw, people can come up with statistics to prove anything, Kent. Forfty percent of all people know that." - Homer Simpson

3.1 Measures of Center

We now enter the Data Analysis and Statistics portion of the course. Everyday we encounter sets of data, from food labels, newspaper articles, etc. It is incredibly important to be able to analyze these data and be able to make judgements based on them.

Probably the most common way to analyze sets of data is to find the **average**. As I'm sure you know, there are typically three different definitions of average: mean, median, and mode.

1. The **mean** of a data set is the sum of all the data points divided by the total number of data points.
2. The **median** of a data set is the middle number of the data set when it is placed in order. If there is an even number of data points, then we typically take the mean of the two middle points.
3. The **mode** of a data set is the set that comes up most often. Note that this need not be unique: you can have two or more data points which both come up the same number of times. In these cases, there is simply more than one mode for a data set.

With these definitions in mind, work on the following problems:

For this problem, you have three sets of data points, taken from students' scores on an algebra quiz (out of 10 points):

$$\text{Class A: } \{4, 8, 4, 3, 4, 0, 6, 10, 4, 6\}$$
$$\text{Class B: } \{9, 0, 2, 1, 6, 10, 1, 6, 5, 9\}$$
$$\text{Class C: } \{7, 6, 8, 2, 2, 4, 5, 3, 8, 5\}$$

1. Before doing any calculations, look at the data sets. What similarities do they have? What differences? Do you think one class did "better" on the quiz than another class? Why?

2. How might you represent these scores visually to demonstrate your answer in part 1)? Come up with more than one way. What are the advantages and disadvantages of each of the ways? Suppose instead of the scores being out of 10, the scores were out of 100. What would be the advantages and disadvantages of your methods?

3. Calculate the mean, median, and mode for each of the data sets. How do these values align with your answers for the first two parts of this question?

4. What exactly does the mean tell you about each data set? For example, if you had a student ask you "Why do we care about the mean?", how would you answer the student? Try to come up with a visual way of explaining the mean.

5. What does the median tell you about each data set? If a student asked "Why do we care about the median, the mean tells us all we need?", how would you respond? Try to come up with a picture to help.

6. What does the mode tell you about each data set? If a student asked "When is the mode ever useful?", how might you respond? Can you come up with an example where the mode tells you something that the mean and median don't?

7. Come up with a set of 5 numbers, so that the median is below 50 and the mean is greater than 200. Which (if any) measure do you think is more representative of your data set?

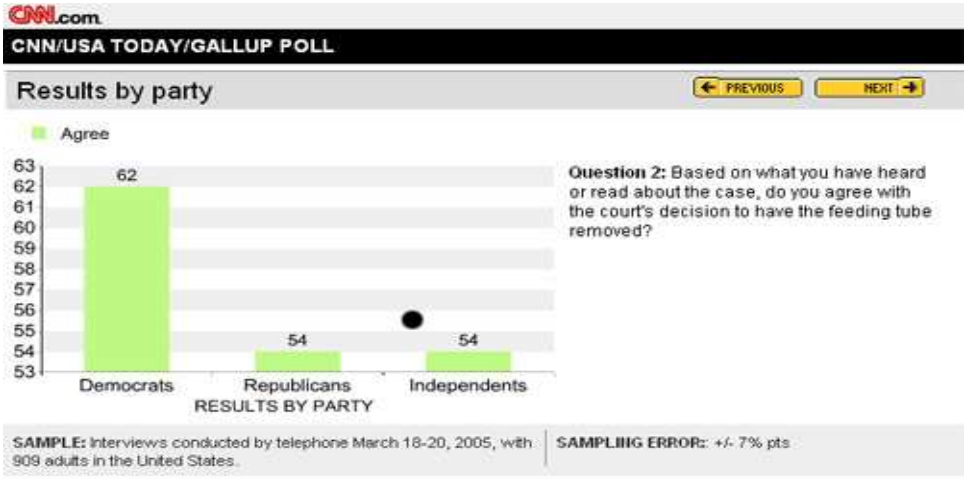
8. Create a set of 5 test scores (so that each number has a value from 0 to 100), and the difference between the mean and the median is bigger than 20. Try to make the difference between the mean and median as big as possible. Which (if any) measure do you think is more representative of your data set?

9. Suppose you know that the mean of a data set is much larger than the median. Can you conclude anything about the relationship between the data points and the mean (i.e. are most data points greater than the mean, or less than the mean)? What if the mean is much smaller than the median?

3.2 Representing Data

As Benjamin Disraeli[1] once said, "There are three kinds of lies: lies, damn lies, and statistics." Here, we examine several presentations of statistics which are misleading.

1. CNN polled 909 adults and asked if they agreed with the court's decision to have Terri Schiavo's feeding tube removed. The results of that poll were posted in this manner on cnn.com.



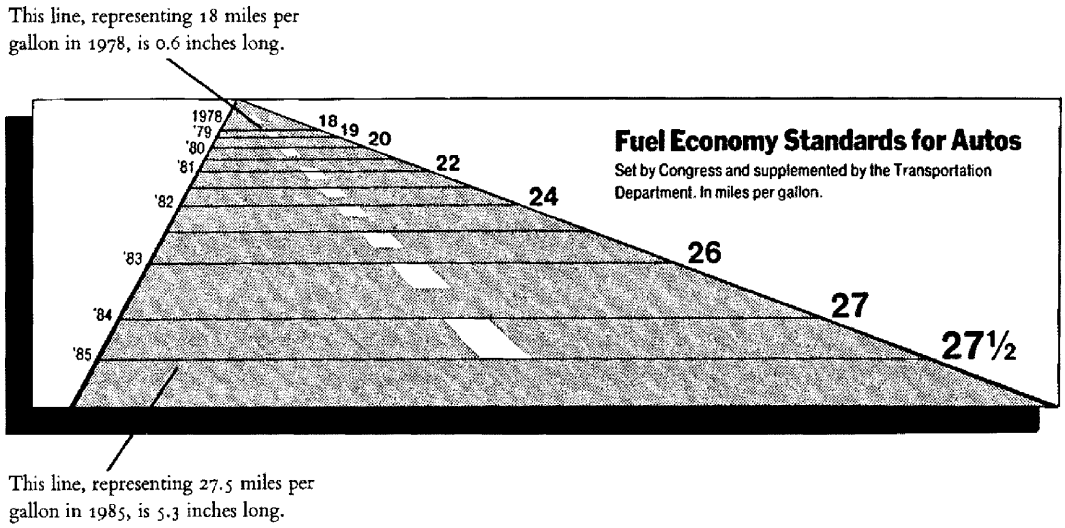
   (a) What are your first impressions upon viewing the graph? What conclusions did you draw?

   (b) How is this graph misleading? How does it misrepresent the actual data?

   (c) Present this data in such a way so that those misrepresentations are no longer present.

[1] A Prime Minister of England during the 19th Century

2. The following graph was printed in the New York Times. It is supposed to represent the fuel economy standards for automobiles as set forth by Congress in 1978.



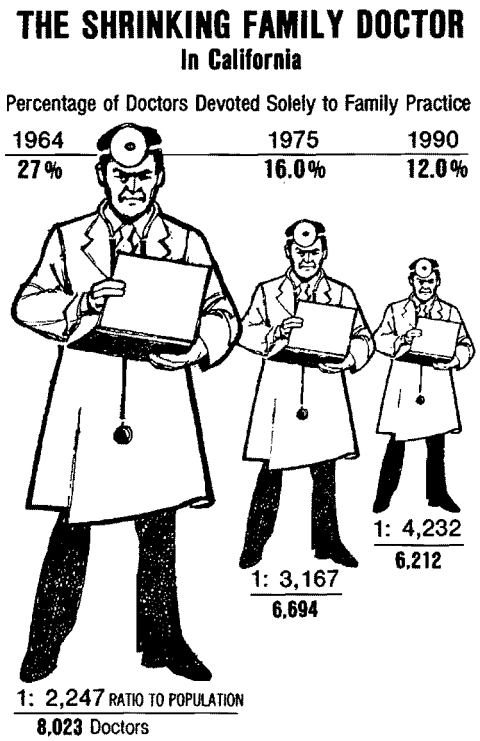
New York Times, August 9, 1978, p. D-2.

(a) Without looking at the numbers, what are your first impressions of the graph? How would you describe the increase in standards (i.e. is it a small increase, large increase, etc.)? How do the years 1978 and 1985 compare?

(b) Now look at the numbers. How does this graph misrepresent those numbers? Using the information provided, explain quantitatively why this graph is misleading.

(c) Again, create your own graph of the data so that these misrepresentations are no longer present.

3. The final graph was taken from the Los Angeles Times.



Los Angeles Times, August 5, 1979, p. 3.

(a) This graph is different from the others we've looked at. In the first graph, the height of the bar represented the data. In the second, the length of the lines represented the data. What represents the data here?

(b) Create your own plot of the data which depicts the exact same data. What is different about your graph than the graph presented (don't just say they are different. What specifically about the two depictions are different?)? What is inaccurate about the graph above?

3.3 Box-and-Whisker Plots

You have undoubtedly seen many different ways to graph sets of data (bar graphs, scatter graphs, etc.). We now discuss the box-and-whisker plot (often called a boxplot). This is a relatively new graph (it was first invented by John Tukey in 1977), but it has gained popularity because it has many useful properties. We first describe how to draw a box plot:

1. Draw a line which represents all the possible data points (i.e. for an exam, the line would go from 0 to 100). Arrange the data points numerically from smallest to greatest. Draw two lines above the number line, one above the lowest data point, and one above the greatest data point. Find the median of the data, and draw a line above your number line to represent the median. The median is often called the **middle quartile**

2. The median has divided the data into two halves. Find the median of the bottom half. This new median is called the **lower quartile**.

3. Find the median of the upper half of data points. This median is called the **upper quartile**.

4. Draw two lines connecting the lower quartile to the upper quartile. You should have a box which contains the (overall) median. This is the box.

5. Draw a line off the box connecting to the lowest data point, and a line to the greatest data point. These are the whiskers.

Many times people will not draw a complete number line, but shorten it to a couple of points above and below the values which occur. For example, if you have a set of exam scores, graded out of 100, and everyone gets between 60 and 90, you might make box-and-whisker plots, for example, starting at 55 and ending at 95. There are further variations on this, but we will discuss this basic method.

As we shall see, this graph has some useful qualities.

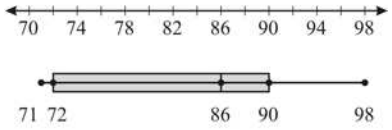
1. Draw box-and-whisker plots for the following sets of data (that is, one box-and-whisker plot for each set of data:

   (a) For quiz scores (out of 10): {2, 5, 5, 2, 4, 7, 10, 3, 9, 9, 5, 8, 4, 7, 9, 5, 9, 6, 9, 10, 10}

(b) For exam scores (out of 100): {47, 39, 52, 73, 67, 89, 97, 54, 60, 82, 91, 76, 53, 58, 77, 78, 59, 43, 74, 96, 93}

(c) For exam scores (out of 200): {146, 180, 162, 164, 62, 105, 147, 160, 26, 152, 150, 38, 39, 159, 125, 155, 88, 114, 145}.

2. Look at the following box-and-whisker plot, for a set of exam scores. What does this tell you about the data? What advantages does this method of graphing have you that other kinds do not?



3.4 Data Collection and Analysis

The situations below depict some of the ways in which data or statistics may be used to reach conclusions. Often the information given will only be part of the picture. Unless otherwise stated, the data below are fictitious.

1. A recent study reported that the number of women working in upper management jobs has increased since 1985.

   (a) What additional information would you need to make a claim about gender equality in the workplace?

   (b) What kinds of claims would this information allow you to make?

2. The percentage of children diagnosed with autism has increased dramatically over the last ten years.

   (a) Come up with at least two interpretations of these results.

   (b) What additional information would help you reach a conclusion?[2]

3. A carefully designed study shows a highly significant correlation between ethnicity and performance on the NAEP (the National Assessment of Educational Progress, or "the Nation's Report Card").[3]

   (a) Come up with three interpretations of these results.

   (b) How could you modify the study to get to reach a conclusion?

4. A study was conducted on the success of a new curriculum. The study took students whose parents signed them up for the new curriculum, and a matching control group of students who did not sign up and who were as similar as possible to the first group (age/grade, performance on standardized tests, grade point average, ethnicity, gender, etc.). At the end of the year the students that completed the program did better on a standardized test. Assume that there were a lot of students in the study (enough to make sure that it was a large enough sample).

   (a) Come up with at least three interpretations of these results.

   (b) What additional information would you need to reach a conclusion? If needed, how could you modify this study to get this information?

5. A researcher is interested in the severity of the crimes committed by African Americans versus whites. She has found that African Americans get significantly longer sentences.

[2]For one study on this topic see `http://www.bmj.com/cgi/content/full/328/7436/364-b`

[3]You can find results from the NAEP at `http://nces.ed.gov/nationsreportcard/`

(a) Can you use this information to make an argument about the severity of the crimes African Americans commit? What are at least two interpretations of this data?

(b) What further information would you need to reach a conclusion?

Themes

The point of the above scenarios is not that statistics and research is unreliable, but that you have be critical about understanding the details of the study. Below are three themes that I came up with as things to think about when interpreting a study, but there are certainly other ways to think about this. Discuss each theme in your group and comment briefly on how the theme shows up in each of the five scenarios above.

- What question are you asking and how are you measuring that?
- How are you comparing different groups?
- What conclusions are you reaching about causality?