**Name:**

**Classwork 1.2 (Week 1-Thursday)**

**Today’s Schedule:**

* “You can grow your brain” article
* Discrimination or Random Chance?
  + Class part
  + Lab part

**You can grow your brain**

Read through the article.

1. Imagine someone who says, “I’m just not a math person. I used to be good at math but now it just doesn’t make any sense to me.” Based on the article, what is something you can say to encourage this person?
2. We now know that Dr. Ji has high expectations. According to the article, why is her class designed in this crazy way?

**Discrimination or Random Chance?**

**CLASS PART**

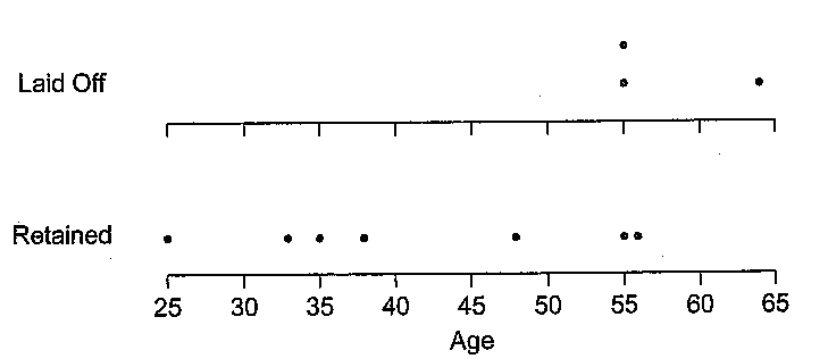
In the year Robert Martin turned 54, the Westvaco Corporation decided to downsize. They laid off more than half of the engineering department, including Robert Martin. Later that year, he sued Westvaco, claiming he had been laid off because of his age. A major piece of Martin’s case was based on a statistical analysis of the ages of the Westvaco employees.

This case is not about age; it is about discrimination. It is about an employee fighting back against what he sees as unfair treatment. If you are fired from a job, maybe you weren’t very good at it or maybe you were just unlucky.

On the other hand maybe it was because someone thought you were the “wrong” gender, or your skin was the “wrong” color, or you looked too young or too old. *How do you know?* While the use of statistics alone cannot *prove* discrimination, statistics can provide evidence by detecting patterns that are consistent with the practice of discrimination.

Let’s look at one department of the Westvaco Co. They had 10 hourly workers. Their ages arranged from youngest to oldest were 25, 33, 35, 38, 48, 55, 55, 55, 56, 64. The three workers who were laid off were ages 55, 55, and 64.

We can visualize these two distributions of people, those that were retained and those that were laid off.



We can also summarize these distributions. For instance, we could “condense” the data into a single number called a “summary statistic.” One possible summary statistic is the average, or mean, age of the three workers who lost their jobs.

1. Take a moment to practice putting MEANING onto numbers. What does this 58 mean? What does it stand for?

Summary of three people who got fired; “middle” number

Key point: one number that stands for three people who got fired

1. Check out the data but be careful not to jump to conclusions. What’s your opinion about the Westvaco data? Do you suspect discrimination? Consider both sides.
   1. What are some reasons (from the data) you have for suspecting discrimination?

The three people were all from the high end of the ages in this department.

Oldest person got fired.

No “young” people were fired.

* 1. What are some reasons (from the data) you have for not suspecting discrimination?

There were some similar age people who did not get fired (55, 56).

Half of the department was over 50 to begin with.

* 1. Is it possible to have gotten this pattern of data by firing three people at random?

Yes.

Take a moment to read this dialogue out loud with a partner next to you.

|  |  |
| --- | --- |
| Martin sympathizer | Look at the pattern in the data. All three workers laid off were much older than the average age of the retained workers. 58 vs. 41.4. That is clear evidence of age discrimination. |
| Westvaco sympathizer | You know it is totally possible to have just fired three people at random and to have their average age be 58. Besides, you are only looking at 10 workers total and only three positions were eliminated. Just one small change and the picture would be totally different. |
| Martin  sympathizer | What do you mean? |
| Westvaco  sympathizer | Imagine just three people were randomly fired. The 25-year-old could have been fired instead of the 64-year-old.  Actual data: 25 33 35 38 48 ~~55~~ ~~55~~ 55 56 ~~64~~  Imagined data: ~~25~~ 33 35 38 48 ~~55~~ ~~55~~ 55 56 64  See! Just one small change and now the average age of those that were fired (45) is lower than the average age of those that were retained (47). |
| Martin  sympathizer | Of all the possible changes, you picked the one most favorable to your side. Some substitutions would have made the averages look even worse! For instance, if you kept one of the 55-year-olds but fired the 56-year-old. Why not compare what actually happened with ALL of other possibilities. |
| Westvaco sympathizer | What do you mean? |
| Martin  sympathizer | Start with 10 workers and pick three at random. Do this over and over to see what typically happens and compare the actual data with those hypothetical data. |
| Westvaco sympathizer | Fine, let’s do it then. |

1. If you pick three of the ten ages [25, 33, 35, 38, 48, 55, 55, 55, 56, 64] at random, do you think you are likely to get an average age of 58 or greater? Why or why not?

(Common student explanation for “Likely”) Well, many of the ages are older than 50.

(Common student explanation for “Not likely”)

\*Only way to get an average age of 58 or greater is to fire the 64 year old – only person older than 58.

(Less common explanation)

\*And if you fire the 64 year old but fire anyone younger than 55, you will get an average age younger than 58.

1. If the probability of getting an average age of 58 or greater turns out to be small, does this favor Martin or Westvaco? Why or why not?

(This is a harder question than you would think for students)

This favors Martin. If it was “easy” to fire people with such a high average age, that would have favored Westvaco. But since it isn’t so “easy” to fire three people with an average age of 58, it makes us suspicious – how did they pick these three people to fire? Not a slam dunk on age discrimination, but it puts more burden on the company to explain their process.

1. As a class, let’s figure out whether it is likely or unlikely to get an average age of 58 years or greater if you choose three workers at random.

* What You’ll Need: Tear up a piece of paper into 10 small pieces. Write the ages of the ten workers in this department (one age on each piece of paper). KEEP THESE PIECES OF PAPER FOR LAB SECTION.

1. *Create a model of a chance process****.*** Draw out three (the ones to be laid off), and record the ages here.

\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_

1. *Compute a summary statistic.* Compute the average of the three numbers in your sample to one decimal place.
2. What percentage of the class randomly drew a sample that had an average greater than 58?

Have students raise their hands if they drew a sample with average age greater than 58. Should be very few students. Should be about a 5% chance. In a class of 25, maybe 1-2 students.

1. *Interpret your results.* What do you conclude from your class’s estimate in step 3)? Does this help Westvaco or make it look suspicious?

Go back to answer 4.

1. Draw the distribution triad that Dr. Ji demonstrates for you in the middle of this sheet. Make it fairly large.

Tell them that this structure is here to help us summarize the most important concepts in the whole class. We won’t understand it all today but we’re setting up the structure to help us better understand not only this Westvaco example but to understand how inferential statistics basically works. Just have them write this in (I often just jot it down at the end of class – no time for much explanation. Later we’ll map on the Westvaco example to it.

**Discrimination or Random Chance?**

**LAB PART (google drive)**

1. Now that we have an intuition for what is going on in the Martin v. Westvaco case, we are going to use technology to help us visualize what is happening more precisely. We are going to call this a **simulation.**

So once again, let’s figure out whether it is likely or unlikely to get an average age of 58 years or greater if you choose three workers at random.

* What You’ll Need: The same 10 pieces of paper you used in class.

1. *Create a model of a chance process****.*** Like in class, mix the ages thoroughly, draw out three (the ones to be laid off), and record the ages in google drive.
2. *Compute a summary statistic.* Compute the average of the three numbers using excel function =average(…)
3. *Repeat the process.* Repeat steps 1 and 2 nine more times.
4. *Display the distribution.* Pool your results with the rest of your class and display the distribution of average ages visually. Roughly sketch what that distribution looks like here.
5. *Estimate the probability*. Count the number of times your class computed an average age of 58 years or greater. Estimate the probability that simply by chance the average age of those chosen would be 58 years or greater.

I have the google drive sheet set up to color in all the 58 or over cells red so you just have to count the red cells and divide by the number of total simulated samples (e.g., 25 x 10 = 250). Should be less than 5%.

1. *Interpret your results.* What do you conclude from your class’s estimate in step 5)?

Some people didn’t think this was likely, some people will think it is likely. This is good to have this disagreement because it sets up a need for some standard that we can judge “likelihood.”

1. What is the distribution from Step 4) in the simulation called? Why is that distribution important?

This is the sampling distribution of means (SDoM). This is a distribution of a bunch of simulated sample distributions.

This is the distribution that helped us figure out what the probability of getting a sample age 58 or over might be.

1. Why is 58 so special? What distribution does that number represent?

That is the average age of the actual 3 people who got fired; this is the “empirical sample” to be distinguished from our simulated samples.

1. In a typical court case, a probability of .025 or less is required to serve as evidence of discrimination. This is the legal definition of “too unlikely.”
   1. Did the layoffs of workers in this department meet the court requirement? What would the court’s ruling be given our simulation?
   2. If our simulation had shown that the probability in the Martin case had been .01 instead of \_\_\_\_\_\_\_\_\_, what would the court’s ruling be?
   3. If our simulation had shown that the probability in the Martin case had been .1 instead of \_\_\_\_\_\_\_\_\_, what would the court’s ruling be?
2. Why the thinking that we did with the simulation considered inferential statistics rather than descriptive statistics?

We went beyond what we could have learned by just examining the sample. We looked at a lot of possible samples that a data generating process – a random and unbiased one – could have produced. That gave us a background for looking at our empirical sample. That process gave us new information (the probability of getting 58 or greater by random chance) and we can use that to judge whether the initial firing process was an unbiased process (like random chance) or something else (like discrimination!).