**Name:**

**Classwork 2**

**Advanced Learning Objectives:**

* Construct a sampling distribution of means
* Compare a sample distribution to a sampling distribution of means
* Assess the statistical likelihood ofa particular outcome

**Discrimination or Random Chance?**

**CLASS PART**

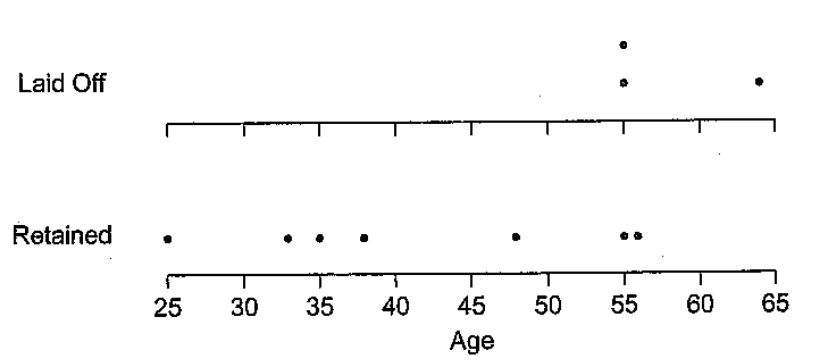
In the year Robert Martin turned 54, the Westvaco Corporation decided to downsize. They laid off more than half of the engineering department, including Robert Martin. Later that year, he sued Westvaco, claiming he had been laid off because of his age. A major piece of Martin’s case was based on a statistical 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 (here, we are simply using DESCRIPTIVE statistics).

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

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?
   2. What are some reasons (from the data) you have for not suspecting discrimination?
   3. Is it possible to have gotten this pattern of data by firing three people at random?

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 years old. 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?

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?

**Discrimination or Random Chance?**

**LAB PART**

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).

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 (we are still in the “DESCRIPTIVE zone,” not yet making inferences).
2. What percentage of the class randomly drew a sample that had an average greater than 58?
3. *Interpret your results.* What do you conclude from your class’s estimate in step 3)? Does this help Westvaco or make it look suspicious?
4. But that’s just 25 hypothetical samples or so. We are going to use technology to help us visualize what is happening more precisely. We are going to call this a **simulation.** We are simulating the chance process of firing three people at random and taking the average age of that sample and plotting the results, over and over again. It can also be called a **sampling distribution**.

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 by constructing a sampling distribution of means.

* What You’ll Need:

1. *Create a model of a chance process****.*** Mix the ages thoroughly, draw out three (the ones to be laid off), and record the ages in the google sheet (**PSY3020 Canvas → Assignments → Labwork 2 Martin v. Westvaco**).
2. *Compute a summary statistic.* Compute the average of the three numbers using excel function =average(…)
3. *Repeat the process.* Repeat until you have 10 simulated samples.
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.e., compare the simulation/sampling distribution of means to the empirical sample of the workers who were actually fired….now we are entering the “INFERENTIAL zone” and going beyond simple descriptions).
6. *Interpret your results.* What do you conclude from your class’s estimate in step 5)? (Assess the statistical likelihood ofthis particular outcome).
7. The distribution from Step 4) in the simulation is called the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_. Why is that distribution important?
8. Why is 58 so special? What distribution does that number represent?
9. In a typical court case, a probability of .025 or less is considered an “unlikely” and if it is unlikely to have turned out like this with a fair process (such as chance), then the court will consider this 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?
10. Why was the thinking that we did with the simulation considered inferential statistics rather than descriptive statistics?

Take Home Messages:

* You can learn anything with effort, good strategies, and help!
* Stats can be used to answer important or interesting questions
* Descriptive statistics: Data can be *described* using visualizations (graphs, etc.) or with summary values (mean, median, mode, standard deviation, etc.)
* Inferential statistics: Sampling Distributions (simulations of repeated sampling) can be compared to our empirical sample to determine the probability of getting a particular outcome by chance (we are not just “describing” the distribution of data, but making inferences about *how likely an outcome is* based on a simulated comparison distribution).