Introduction Activity

Westvaco Discrimination Case

Welcome to AP Statistics! Before we begin our formal study, we will spend a couple days exploring the thinking and concepts we will encounter throughout the course of the year.

An Introduction to Statistics:

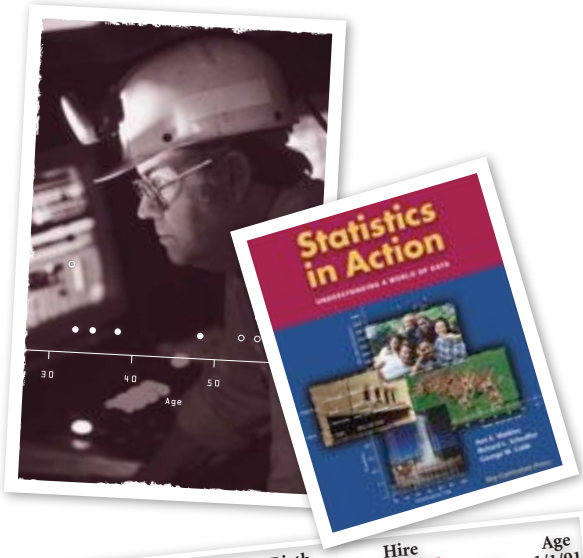
- Statistics vs. Other Mathematics
- Westvaco Discrimination Case Data and Discussion

A Case Study of Statistics in Action

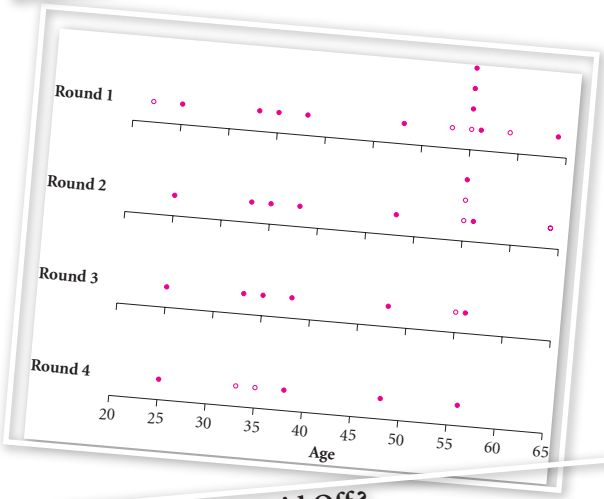
Martin v. Westvaco

{Adapted from Statistics in Action: Watkins, Scheaffer, Cobb.}

Robert Martin turned 55 in 1991. Earlier in that same year, the Westvaco Corporation, which makes paper products, decided to downsize. Several members of their Engineering Department were laid off, including Mr. Martin. Later that year, Martin sued Westvaco, claiming he had been laid off because of his age. A major piece of his case was based on statistical analysis of the ages of employees at Westvaco.



Row	Job Title	Pay	Birth Mo	Birth Yr	Hire Mo	Hire Yr	RIF	Age 1/1/91
1	Engineering Clerk	H	9	66	7	89	0	25
2	Engineering Tech II	H	4	53	8	78	0	38
3	Engineering Tech II	H	10	35	7	65	0	56
4	Secretary to Engin Manag	H	2	43	9	66	0	48
5	Engineering Tech II	H	8	38	9	74	1	53
6	Engineering Tech II	H	8	36	3	60	1	55
7	Engineering Tech II	H	1	32	2	63	1	59
8	Parts Crib Attendant	H	11	69	10	89	1	22
9	Engineering Tech II	H	5	36	4	77	2	55
10	Engineering Tech II	H	8	27	12	51	2	64



Laid Off?

Under 40?	Yes	No	Total	% Yes
Yes	4	5	9	44.4
No	14	13	27	51.9
Total	18	18	36	50.0

In this activity, you will be introduced to and get a chance to try your hand at two types of statistical work:

- **Exploratory Data Analysis**
- **Statistical Inference**

The purpose of this activity is to give you a head start on the ideas of statistical thinking, before you get involved with the details of the methods.

It is easy to get caught in the trap of *doing* rather than *understanding*, of asking *how* rather than *why*. You can't *do* Statistics unless you learn the methods, but you must not get so caught up in the details that you lose sight of what they mean.

Discrimination in the Workplace: Exploratory Data Analysis

At the beginning of 1991, Robert Martin was one of 50 people employed in the engineering department of Westvaco's envelope division. That spring, Westvaco's management underwent five rounds of reductions in their workforce. In Round 1, 11 positions were eliminated. 9 more were eliminated in Round 2. By the time all 5 rounds were completed, 28 of the 50 employees lost their jobs and the average age in the department fell from 48 to 46 years old.

Through this activity, we will explore the following basic ideas of statistics:

- ☑
- ☑

Display 1.1 (on the next page) shows data provided by Westvaco to Martin's lawyers. Each row corresponds to one worker, and each column corresponds to a characteristic of the worker: job title, hourly v. salaried, date of birth, date of hire, and age as of January 1, 1991 (shortly before layoffs). The "RIF" column indicates how the worker fared in the rounds of layoffs. "1" indicates they were laid off in Round 1, "2" in Round 2, etc. "0" indicates they were not laid off.

Define the following in the context of the Westvaco case:

☑ **Cases:**

☑ **Variables:**

☑ **Variability:**

☑ **Distribution:**

Why is **variability** important to the study of Statistics?

Define **Statistics:**

Row	Job Title	Pay	Birth		Hire		RIF	Age 1/1/91
			Mo	Yr	Mo	Yr		
1	Engineering Clerk	H	9	66	7	89	0	25
2	Engineering Tech II	H	4	53	8	78	0	38
3	Engineering Tech II	H	10	35	7	65	0	56
4	Secretary to Engin Manag	H	2	43	9	66	0	48
5	Engineering Tech II	H	8	38	9	74	1	53
6	Engineering Tech II	H	8	36	3	60	1	55
7	Engineering Tech II	H	1	32	2	63	1	59
8	Parts Crib Attendant	H	11	69	10	89	1	22
9	Engineering Tech II	H	5	36	4	77	2	55
10	Engineering Tech II	H	8	27	12	51	2	64
11	Technical Secretary	H	5	36	11	73	2	55
12	Engineering Tech II	H	2	36	4	62	3	55
13	Engineering Tech II	H	9	58	11	76	4	33
14	Engineering Tech II	H	7	56	5	77	4	35
15	Customer Serv Engineer	S	4	30	9	66	0	61
16	Customer Serv Engr Assoc	S	2	62	5	88	0	29
17	Design Engineer	S	12	43	9	67	0	48
18	Design Engineer	S	3	37	6	74	0	54
19	Design Engineer	S	3	36	2	78	0	55
20	Design Engineer	S	1	31	3	67	0	60
21	Engineering Assistant	S	6	60	7	86	0	31
22	Engineering Associate	S	2	57	4	85	0	34
23	Engineering Manager	S	2	32	11	63	0	59
24	Machine Designer	S	9	59	3	90	0	32
25	Packaging Engineer	S	3	38	11	83	0	53
26	Prod Spec—Printing	S	12	44	11	74	0	47
27	Proj Eng—Elec	S	9	43	4	71	0	48
28	Project Engineer	S	7	49	9	73	0	42
29	Project Engineer	S	8	43	4	64	0	48
30	Project Engineer	S	6	34	8	81	0	57
31	Supv Engineering Serv	S	4	54	6	72	0	37
32	Supv Machine Shop	S	11	37	3	64	0	54
33	Chemist	S	8	22	4	54	1	69
34	Design Engineer	S	9	38	12	87	1	53
35	Engineering Associate	S	2	61	9	85	1	30
36	Machine Designer	S	2	39	4	85	1	52
37	Machine Parts Cont—Supv	S	10	28	8	53	1	63
38	Prod Specialist	S	9	27	10	43	1	64
39	Project Engineer	S	7	25	9	59	1	66
40	Chemist	S	12	30	10	52	2	61
41	Design Engineer	S	4	60	5	89	2	31
42	Electrical Engineer	S	11	49	3	86	2	42
43	Machine Designer	S	3	35	12	68	2	56
44	Machine Parts Cont Coor	S	9	37	10	67	2	54
45	VH Prod Specialist	S	5	35	9	55	2	56
46	Printing Coordinator	S	2	41	1	62	3	50
47	Prod Dev Engineer	S	6	59	11	85	3	32
48	Prod Specialist	S	7	32	1	55	4	59
49	VH Prod Specialist	S	3	42	4	62	4	49
50	Engineering Associate	S	8	68	5	89	5	23

Display 1.1 The data in *Martin v. Westvaco.*

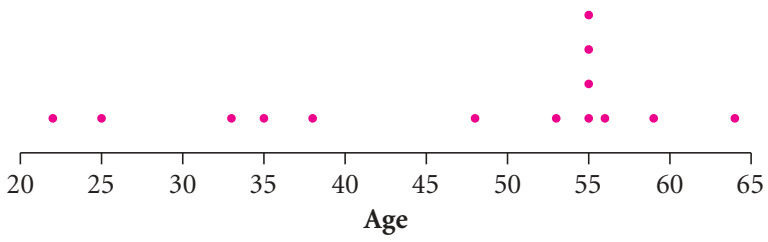
Source: *Martin v. Envelope Division of Westvaco Corp.,* CA No. 92-03121-MAP, 850 Fed. Supp. 83 (1994).

D1: If you were on the jury in the Martin v. Westvaco Case, how would you use the information in Display 1.1 to decide if Westvaco tended to lay off older workers?

Throughout the course of our study, we will rely on a variety of displays to explore distributions of data. The distribution of data provides us with information on the values a particular variable takes on, how often it takes on those values, how spread out they are, and whether or not any unusual values are present. When exploring data, it helps to make a display of the distribution and discuss the "SOCS":

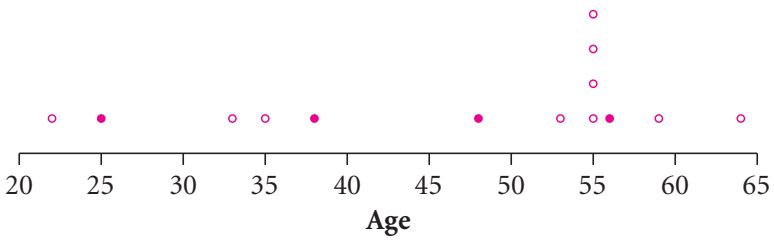
- **S**
- **O**
- **C**
- **S**

Dotplot of Ages of Hourly Workers



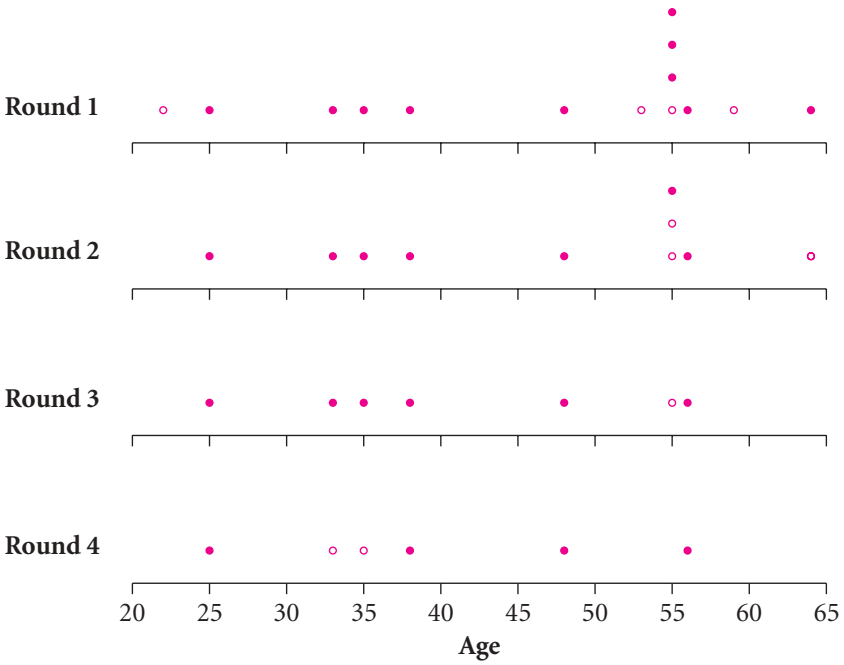
Display 1.2 Ages of the hourly workers. (Each dot is a worker; the ages are shown by the position of the dots along the scale below them.)

Dotplot of Ages of Hourly Workers, Laid-off vs. Retained



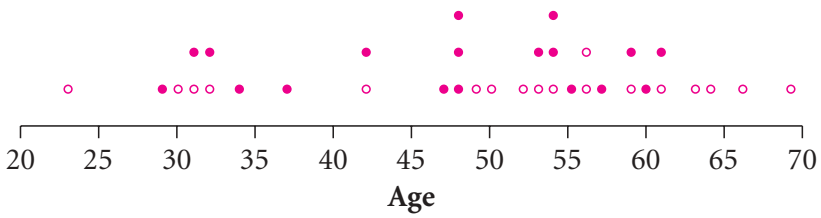
Display 1.3 Hourly workers: Ages of those laid off (open circles) and those retained (solid dots).

D2: Does the dotplot in Display 1.3 show a "clear cut" case of age discrimination, a possible case of discrimination, or no discrimination? Why or why not?



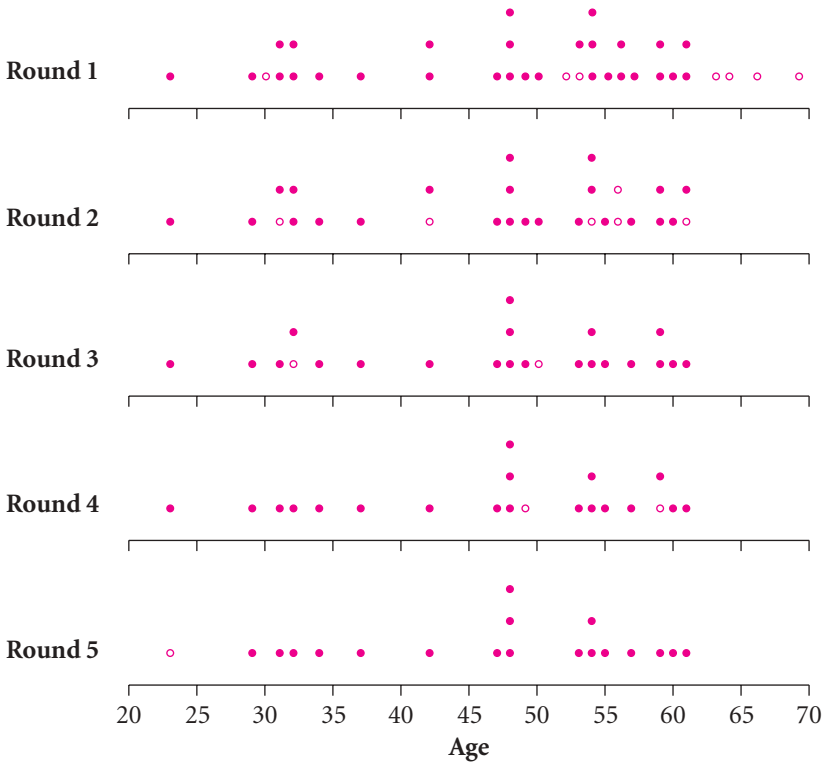
Display 1.4 Hourly workers: Ages of those laid off (open circles) and those retained (solid dots) in each round.

D3: The dotplots in Display 1.4 show the ages of the hourly workers laid off and retained by round. Compare the round-by-round information in Display 1.4 with the summary for all rounds in Display 1.3. Which display provides a stronger case for discrimination? Why?



Display 1.5 Salaried workers: Ages of those laid off (open circles) and those retained (solid dots).

D4: The dotplot in Display 1.5 is similar to Display 1.3, except that it includes salaried workers. Compare the plots for the hourly and salaried workers. Which provides stronger evidence in support of a claim for age discrimination? Why?



Display 1.6 Salaried workers: Ages of those laid off (open circles) and those retained (solid dots) in each round.

D5: Display 1.6 is similar to Display 1.4. It shows the distributions of ages and layoffs for salaried workers. Compare the pattern of salaried workers with the pattern for hourly workers in Display 1.4.

Summary of Salaried Workers - Age and Employment Status

		Laid Off?			
		Yes	No	Total	% Yes
Under 40?	Yes	4	5	9	44.4
	No	14	13	27	51.9
	Total	18	18	36	50.0

D6: The summary table shown here classifies salaried workers using two Yes-No questions: Under 40? and Laid Off?

a. Does the pattern in this table support Martin's claim of discrimination? Why or Why not?

b. Construct a similar table for salaried workers, using 50 to divide the ages. Does the evidence in this new table provide stronger or weaker support for Martin's case? Explain. Compare both.

		Laid Off?			
		Yes	No	Total	% Yes
Under 50?	Yes				
	No				
	Total				

D7: Whenever you think you have a message from data, you should be careful not to jump to conclusions. The patterns in the Westvaco data might be "real" - meaning they reflect age discrimination. On the other hand, the patterns might be the result of "chance" - management wasn't discriminating on the basis of age, but simply happened to lay off a larger percentage of older workers.

D8: You may feel the analysis so far ignores important facts like worker qualifications. That's true...however, the first step is to decide if, based on the data in Display 1.1, older workers were more likely to be laid off. If not, Martin's case fails. If so, however, it is up to Westvaco to justify its actions.

Practice:
Construct a dotplot similar to Display 1.3, comparing the ages of hourly workers who lost their jobs at some point during the first 3 rounds to the ages of hourly workers who still had their jobs at the end of round 3. How do the ages differ?

In D6, you compared summary tables for salaried workers. Construct similar tables for the hourly workers, using 40 as a cutoff for one and 50 as a cutoff for the other. Which table provides stronger evidence for discrimination? How do the patterns compare to the salaried workers?

	Laid Off?				
Under 40?		Yes	No	Total	% Yes
	Yes				
	No				
	Total				

	Laid Off?				
Under 50?		Yes	No	Total	% Yes
	Yes				
	No				
	Total				

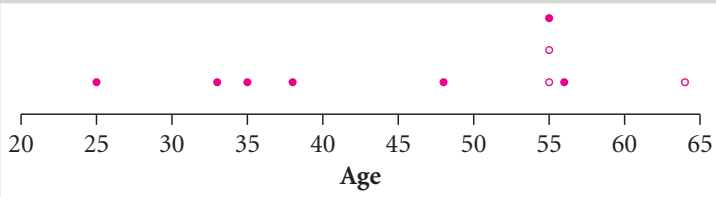
Discrimination in the Workplace: Statistical Inference

The exploratory data analysis of the Westvaco case suggests that older workers were more likely to be laid off than younger ones. One of the main arguments in the case dealt with what those patterns mean:

- Can we infer that Westvaco has some explaining to do?
- Could the patterns happen if there was no discrimination?

We can't answer those questions completely--we just don't know enough yet. However, you can still get a pretty good idea of how the analysis works by considering a subset of the data.

Consider the ten hourly workers involved in Round 2. Their ages were 25, 33, 35, 38, 48, 55, 55, 55, 56, and 64. The three who were laid off were 55, 55, and 64.



Display 1.7 Hourly workers: Ages of those laid off (open circles) and those retained (solid dots) in Round 2.

To simplify the statistical analysis, it helps to "condense" the data into a single number, or **summary statistic**, such as the **mean**. Calculate the average age of the three who lost their jobs:

average = mean = =

D9: This suggests older workers may have been more likely to be laid off...however, could we get an average age of 58 if we picked three of these workers at random? How likely do you think it is that that would happen?

D10: If the probability of getting an average age of 58 or more at random turns out to be small, does that favor Martin or Westvaco?

Discrimination in the Workplace: Statistical Inference Activity

Let's see what happens when we randomly lay-off three workers from Round 2.

- Get 10 index cards from your teacher. Write each of the ten ages of the workers in Round 2 on the cards and mix them thoroughly.
- Draw 3 cards and record the ages.
- Calculate the average of the three ages in your *sample* to one decimal place.
- Repeat the process until you have 10 sample averages.
- Display the distribution of average ages in a summary dotplot.
- Estimate the probability: Count how many average ages were 58 or older and divide by the total number of simulated trials.
- Interpret your results...is it likely that Westvaco could have randomly selected 3 workers with an average age of 58 or older?

Trial	Age 1	Age 2	Age 3	Average Age
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				

Consider the Class Results:

Total Number of Sample Averages ≥ 58: _____

Total Number of Trials: _____

Probability of Randomly Selecting 3 workers with an average age ≥ 58:_____

Inference is a statistical procedure that involves deciding whether or not an event can reasonably be attributed to chance or whether you should investigate some other explanation.

D11: Why did we estimate the probability of getting an average age greater than or equal to 58 rather than just an average of 58?

D12: How unlikely is "too unlikely"? At what point would you be convinced that something other than chance behavior was responsible for the selection of workers?

D13: Suppose a friend wants to bet with you on the outcome of a coin toss. To check the fairness of his coin, you flip it 20 times--it comes up heads 19 times. Why does this evidence make it hard to believe the coin is fair? How could you simulate this situation to estimate how unlikely this result is?

In this course, we will explore the methods of collecting, analyzing, interpreting, and making inferences from data. This case study illustrates just some of the thinking you will be expected to carry out. Many of the methods require little mathematical manipulation. Rather, it will be up to you to be able to explain what you observe in the context of a given situation.

My hope is that you will not only perform well on the AP examination, but (more importantly) that you will leave this course with a better understanding of how Statistics can be used in the "real world". I look forward to working with you this year...now let's start digging in!