

Math 140

Introductory Statistics

Professor Silvia Fernández

Lecture 1

Based on the book *Statistics in Action*
by A. Watkins, R. Scheaffer, and G. Cobb.

Statistics

The science of learning from data
in the presence of variability.

Our first problem

Discrimination in the Workplace: Data Exploration

Robert Martin was one of 50 people working in the engineering department of Westvaco's envelope division. One spring, Westvaco's management went through five rounds of planning for a reduction in their work force. In Round 1, they eliminated 11 positions, and they eliminated 9 more in Round 2. By the time the layoffs ended, after all five rounds, only 22 of the 50 workers had kept their jobs. The average age in the department had fallen from 48 to 46.

After Martin, age 54, was laid off, he sued Westvaco for age discrimination.

The data in Martin v. Westvaco.

Row	Job Title	Pay	Birth		Hire		Round	Age at Birthday in 1991
			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

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

Statistical Work

- Data Exploration
 - Examination of data for *patterns*.
 - Tools: summary tables, graphs, averages, etc.
- Inference (making inferences from data)
 - Definition: Deciding whether or not an observed feature of the data could reasonably be attributed to chance.

Data from Tables

Variables [columns]
 Characteristics of each case.
 It allows us to see the **variability**

Cases [rows]
 Subjects/objects
 of statistical examination

Row	Job Title	Pay	...	Round	Age
2	Engineering Clerk	H		0	25
3	Engineering Tech I	H		0	38
4	Engineering Tech II	H		0	56
...					
...					

In the example:

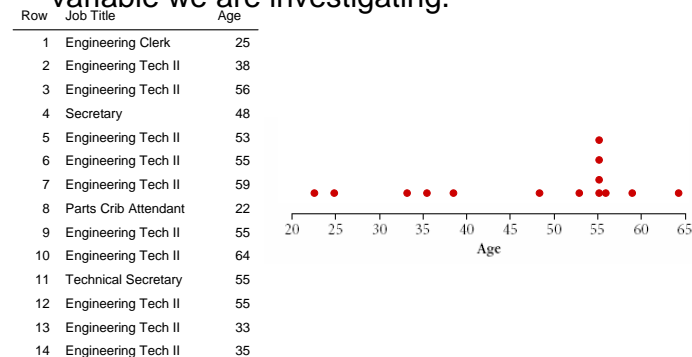
- **Cases** = individual Westvaco employees
- **Variables** = year of birth, job title, pay, etc.

Understanding Variability

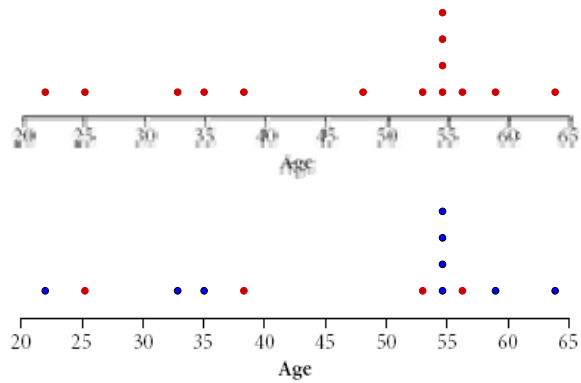
- To understand how the characteristics of the cases varies we look at their **distribution**.
- **Distribution**: What the values are and how often they occur (record of variability)
- How can we study the distribution?
 - By observing the values in each column of the table.
 - By graphing the values in a **dot plot**.

Dot Plots

- Each case is represented by a dot located according to the numerical value of the **variable we are investigating**.



Comparing dot plots



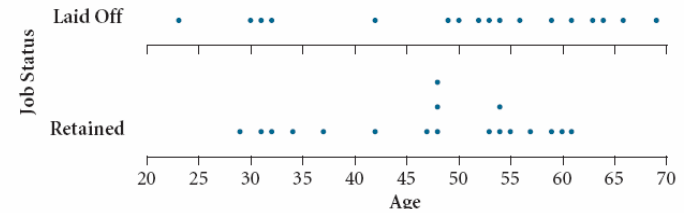
Display 1.3 Hourly workers: Ages of those laid off (open circles) and those retained (solid dots).

Discussion: Exploring the *Martin v. Westvaco* Data

- D1. Suppose you were on a jury in the *Martin v. Westvaco* case. How would you use the information in Display 1.1 ([The table](#)) to decide if Westvaco tended to lay off older workers (for whatever reason)?

Discussion: Exploring the *Martin v. Westvaco* Data

- D2. Compare the plots for the hourly and salaried workers. Which provides stronger evidence in support of Martin's claim of age discrimination?



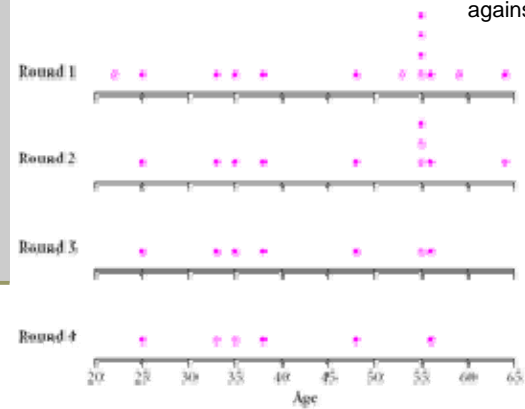
Display 1.5 Hourly workers: ages of those laid off and those retained.

Discussion: Exploring the *Martin v. Westvaco* Data

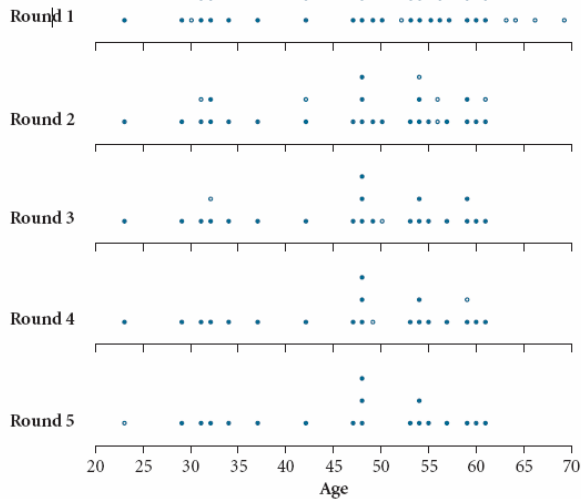
- D3. Whenever you think you have a message from data, you should be careful not to jump to conclusions. The patterns in the Westvaco data might be “real”—they reflect age discrimination on the part of management. On the other hand, the patterns might be the result of chance—management wasn’t discriminating on the basis of age but simply by chance happened to lay off a larger percentage of older workers. What’s your opinion about the Westvaco data: Do the patterns seem “real”—too strong to be explained by chance?
- D4. The analysis up to this point ignores important facts such as worker qualifications. Suppose Martin makes a convincing case that older workers were more likely to be laid off. It is then up to Westvaco to justify its actions. List several specific reasons Westvaco might give to justify laying off a disproportionate number of older workers.

Round by Round

Which display provides stronger support for Martin’s claim that Westvaco discriminated against older workers?



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



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

Using Tables to Compare

- The summary table shown here classifies salaried workers using two yes/no questions: Under 40? and Laid off? (In employment law, 40 is a special age because only those 40 or older belong to what is called the “protected class,” the group covered by the law against age discrimination.)

		Laid Off?		Total	% Yes
		Yes	No		
Under 40?	Yes	4	5	9	44.4
	No	14	13	27	51.9
Total		18	18	36	50

Using Tables to Compare

		Laid Off?			% Yes
		Yes	No	Total	
Under 40?	Yes	4	5	9	44.4
	No	14	13	27	51.9
	Total	18	18	36	50

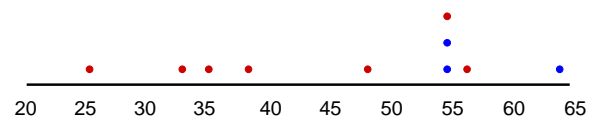
- a. Does the pattern in this table support Martin's claim of age discrimination? Why or why not?
- b. Construct a similar table for salaried workers, but this time use 50 instead of 40 to divide the ages. (Your two age groups will be those under 50 and those 50 or older.) Does the evidence in this new table provide stronger or weaker support for Martin's case? Explain.

How to Analyze Patterns?

- Overall, the exploratory work we just did shows that older workers were more likely than younger ones to be laid off, and they were laid off earlier. One of the main arguments in the court case was about what those patterns mean:
 - Can we infer from them that Westvaco has some explaining to do?
 - Or are the patterns of the sort that might happen even if there was no discrimination?

Summary Statistic

- Consider as an example of our analysis Round 2 of the layoffs.



- To simplify the statistical analysis to come, it will help to "condense" the data into a single number, called a **summary statistic**. One possible summary statistic is the average, or mean, age of the three who lost their jobs:

$$\text{average} = \frac{55 + 55 + 64}{3} = 58 \text{ years}$$