Unit 4: Measures of Center

SUMMARY OF VIDEO

One number most people pay a lot of attention to is the one on their paycheck! Today's workforce is pretty evenly split between men and women, but is the salary distribution for women the same as for men? The histograms in Figure 4.1 show the weekly wages for Americans in 2011, separated by gender.

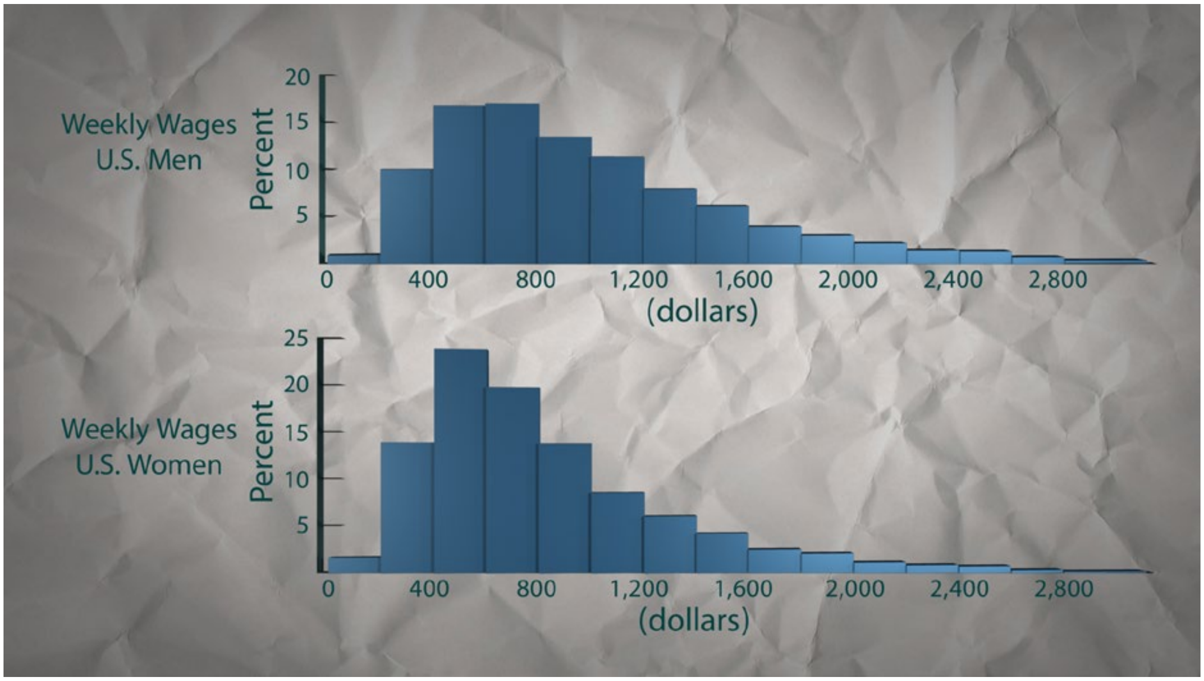

Figure 4.1. Histograms comparing men's and women's wages.

Both histograms are skewed to the right with most people making moderate salaries while a few make much more. For comparison's sake, it would help to numerically describe the centers of these distributions. A statistic called the median splits the distribution in half as shown in Figure 4.2 – half the wages lie above it, and half below. The median wage for men in 2011 was $865. The median wage for women was only $692, about 80% of what men make.

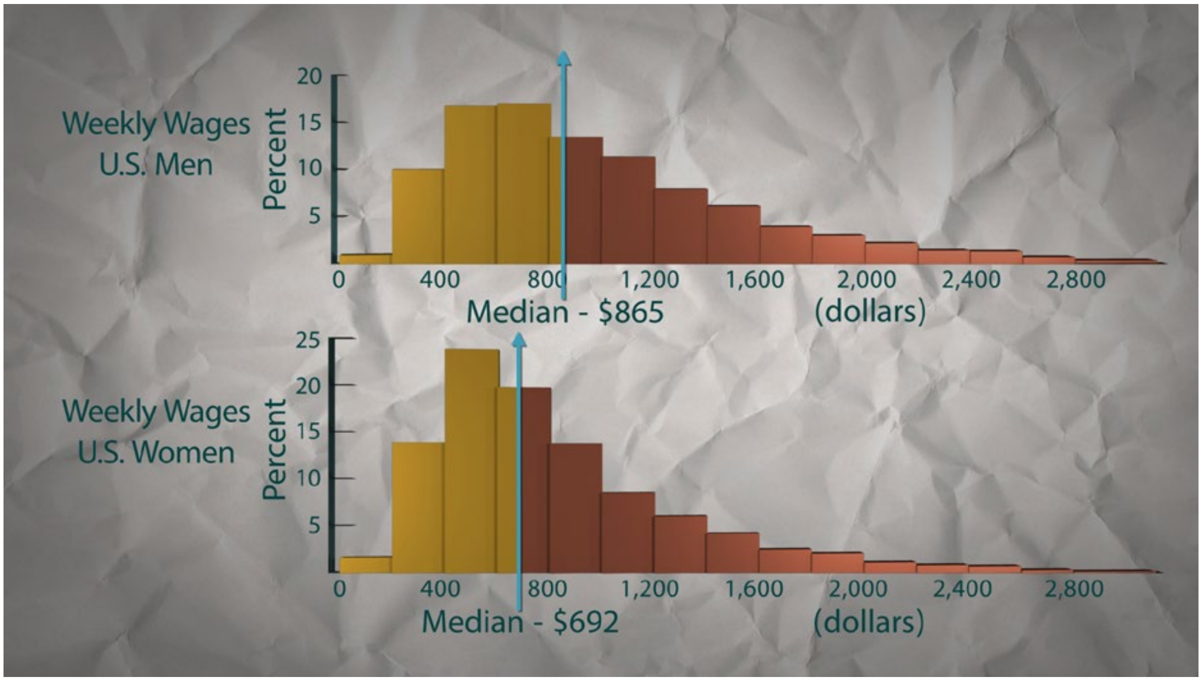

Figure 4.2. Locating the median on histograms of wages.

Simply using the median, we have identified a real disparity in wages, but it is much harder to figure out why it exists. Some of the difference can be accounted for by differences in education, age, and years in the workforce. Another reason for the earnings gap is that women tend to be concentrated in lower-paying jobs – but that begs the question: Are these jobs worth less in some sense? Or are these jobs lower paid because they are primarily held by women? This is the central issue in the debate over comparable worth – the idea that men and women should be paid equally, not only for the same job but for different jobs of equal worth.

Back in 1988 the city of Colorado Springs, Colorado, was at the forefront of this debate. As part of its normal operation, the city government evaluated all municipal jobs with criteria like working conditions, necessary skills, and accountability required. Each job got a numerical rank. It turned out that many clerical jobs, which are mostly filled by women, scored the same number of points as operations and maintenance jobs, which are mostly filled by men. However, the median wage for men's jobs was always higher than the corresponding median wage for the women even though these jobs were judged to be exactly equal in requirements and responsibility. A group of clerical workers used this evidence to pressure the city for a more equitable pay structure. The numbers were hard to argue with and the clerical workers won. The city agreed to equalize the median salaries for jobs of comparable worth. And the plan had a benefit for the city as well – the relatively high turnover rate for jobs held by women decreased.

Colorado Springs relied on the median statistic to identify the inequality in men's and women's salaries. Next, we take a look at how to calculate this measure of center. Below are the weekly salaries from a small hypothetical company that has 19 employees. The salaries have been arranged in order from the lowest, $290 for an entry-level receptionist, up to the highest, $2,000 for the president.

290	350	400	400	450	450	450	500	500	500
550	550	650	750	800	1200	1300	1500	2000	

The median represents a typical wage. To calculate the median, determine the number of observations, n. In this case, we have 19 salaries and so, $n = 19$. The location of the median is at $(n + 1)/2$, or $(19 + 1)/2 = 10$. Count up 10 spots from the bottom (or down 10 spots from the top) and read off the median: $500. Since we had an odd number of paychecks, it is easy to count up 10 places to the middle number. But what if we had an even number of observations to deal with? Suppose we add a paycheck of $550 as shown below.

290	350	400	400	450	450	450	500	500	500
550	550	550	650	750	800	1200	1300	1500	2000

With the additional paycheck, $n = 20$. Now, we count up $(20 + 1)/2$, or 10.5 spaces. That puts us right in between the two middle values of 500 and 550. So, the median is actually halfway between those two salaries, $525.

The median is not the only measure for center. Another way to measure the center of a distribution of values is by taking the average. Statisticians call this number the mean, which is denoted by $\bar{x}$. It is calculated by adding up all the values and dividing by the number of values:

$$\bar{x} = \frac{\sum x}{n}$$

If we return to our original 19 paychecks, we find the mean as follows:

$$\bar{x} = \frac{\$290 + \$350 + \ . \ . \ . \ + \$1500 + \$2000}{19} = \frac{13{,}590}{19} \approx \$715.26$$

Notice that the mean, which is about $715, is higher than the median of $500. You can think of the mean as the balancing point of all the values. It is the value of the pivot point shown in Figure 4.3 that balances all the observations.

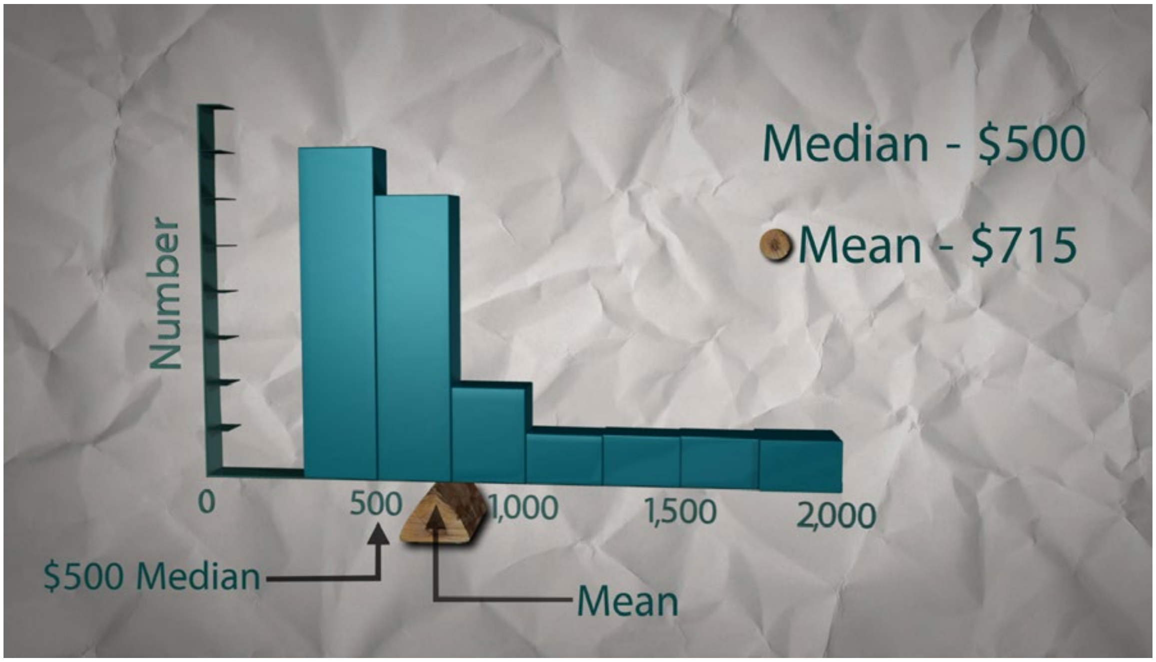

Figure 4.3. The mean as the balancing point.

The mean is influenced much more by the one high salary going to the president of the company. The median, on the other hand, is what statisticians call resistant. The median doesn't depend on what the values are out there at the extremes of our distribution. If the president doubled his salary while everyone else stayed at the same wage, the mean would bump up to $820.53, or around $821. But our median would stay at $500.

The shape of a distribution can give you some hints about the relationship between the mean and median. For a fairly symmetric distribution, such as the one shown in Figure 4.4, the mean and median are roughly the same.

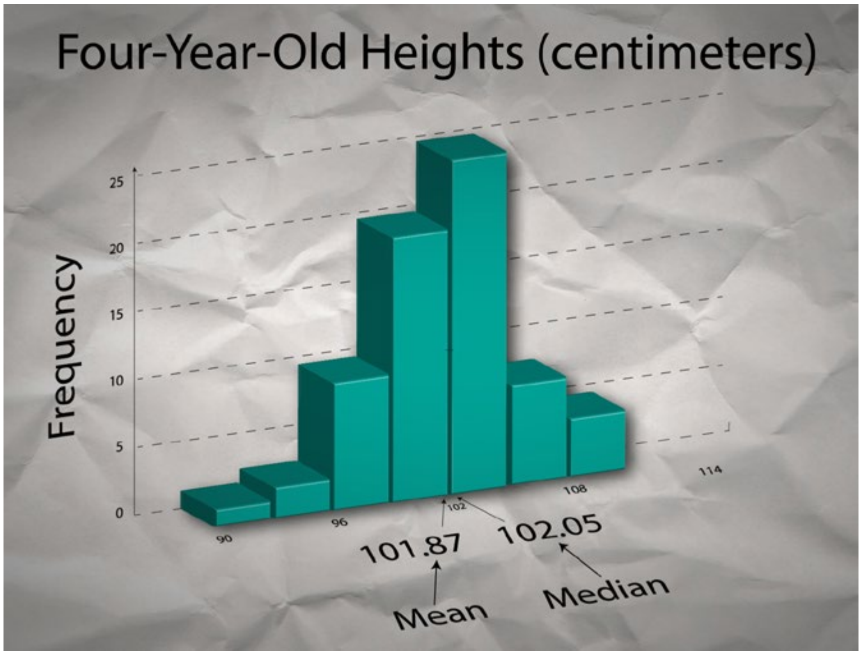

Figure 4.4. Mean and median are close.

If the distribution is skewed to the right, like the scores on the difficult exam pictured in Figure 4.5, the mean is larger than the median.

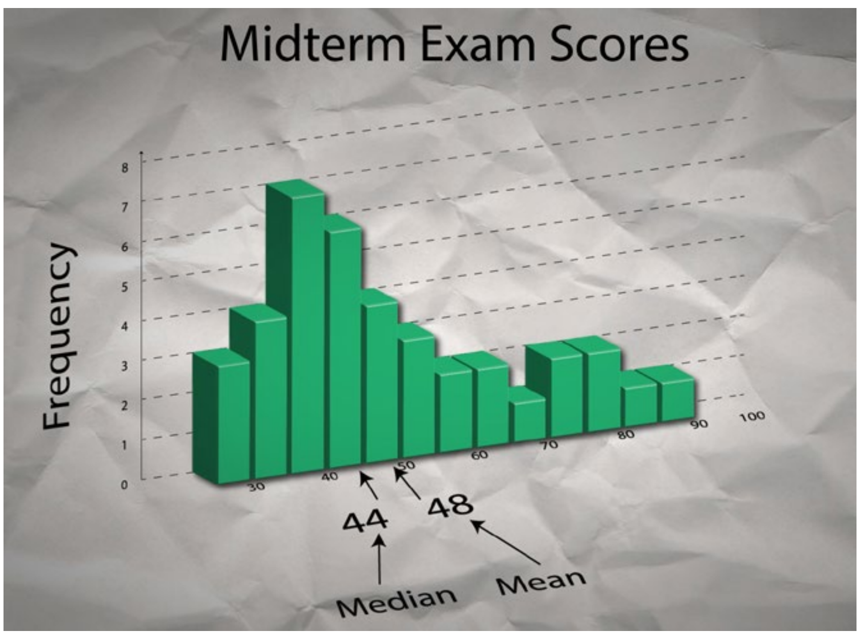

Figure 4.5. Mean larger than median.

Likewise, if the distribution is skewed to the left, like the scores on one easy exam shown in Figure 4.6, the mean is smaller than the median. Remember the mean is influenced by values at the extremes and the median is not.

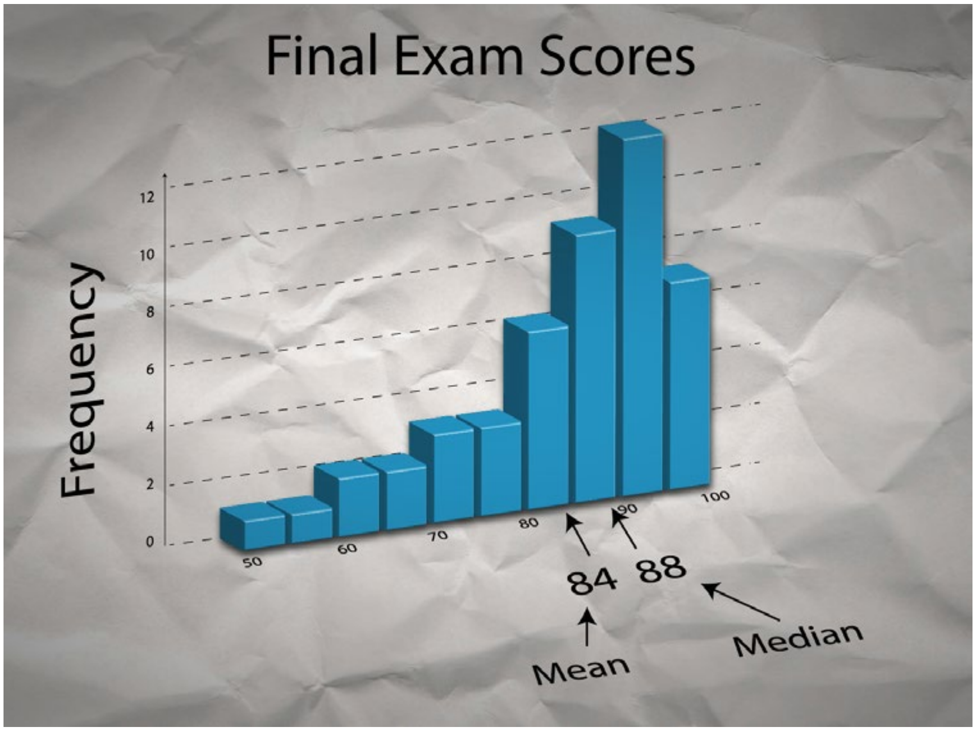

Figure 4.6. Median larger than mean.