

Activity**Describing Distributions Numerically : Income of CEOs***Hand Calculations*

Exercise: Here are the incomes (in millions of dollars) for a sample of 8 CEOs in 2002.

Company	\$
Anheuser-Busch	47
Atmos Energy	3
Dell Computer	16
Eastman Chemical	7
First Tennessee National	5
First Virginia Banks	2
Lone Star Steakhouse	1
Wal-Mart	22

1. Measures of center:

- (a) Finding the mean \bar{y} - pronounced “y-bar”. If there are n observations, then the mean is

$$\bar{y} = \frac{\text{Total}}{n} = \frac{\sum y}{n}$$

The mean is the common arithmetic average. The mean salary for these 8 CEOs is _____.

- (b) Finding the median M .

- i. List all the observations from smallest to largest.
 - ii. If the number of observations is odd, then the median is the middle observation. Count from the bottom of the list of ordered values up to the $(n + 1)/2$ largest observation. This observation is the median.
 - iii. If the number of observations is even, then the median is the average of the two center observations.
- The median salary for these 8 CEOs is _____.

2. Measures of spread:

- (a) The range of the data is defined as the difference between the maximum and minimum values, $\text{Range} = \text{max} - \text{min}$.
- (b) Find the quartiles Q_1 and Q_3
- i. Locate the median.
 - ii. The first quartile, Q_1 , is the median of the lower half of the list of ordered observations.
 - iii. The third quartile, Q_3 , is the median of the upper half of the list of ordered observations.

For the salary data: $Q_1 = \underline{\hspace{2cm}}$ $Q_3 = \underline{\hspace{2cm}}$

- (c) Finding the variance s^2 and the standard deviation s

- i. Take the average (almost) of the squared deviations of each observation from the mean.

$$s^2 = \frac{\sum (y - \bar{y})^2}{n - 1}$$

- ii. The standard deviation is the square root of the variance, i.e., $s = \sqrt{s^2}$.

Observations	Deviations	Squared deviations
y	$y - \bar{y}$	$(y - \bar{y})^2$
47		
3		
16		
7		
5		
2		
1		
22		
sum =		sum =

For the salary data: $s^2 =$ _____ $s =$ _____

3. Find the five-number summary.

CEO Salaries				
Min	Q_1	M	Q_3	Max
4. A boxplot is a graphical display of the five-number summary. Construct the boxplot for the CEO salary data.
5. Notes:
- (a) If you use the mean to describe the center, then the standard deviation and variance are the appropriate measures of spread.
 - (b) If you use the median to describe the center, then the quartiles and the interquartile range ($IQR = Q_3 - Q_1$) are the appropriate measures of spread.
 - (c) The mean and standard deviation can be strongly affected by outliers and are harder to interpret for skewed data.
 - (d) The median and quartiles are more appropriate when outliers are present or when the data are skewed.
 - (e) Call an observation an outlier if it falls more than $1.5 \times IQR$ above the third quartile or below the first quartile.
6. Which measures of center and spread are appropriate for the salary data?
7. Are there any outliers present?
8. What happens to the summary statistics (means, medians, quartiles, standard deviations) if the largest value is removed from the dataset?
9. What happens to the summary statistics if each data value is increased by 1?
10. What happens to the summary statistics if each data value is multiplied by 1 million?