Lecture 5: Descriptive Statistics

Mean, Median, Mode, Standard Deviation

Tom Hanna

2026-02-05

Agenda

- Lecture today: Descriptive statistics

```
  - Measures of Central Tendency: Mean, Median, Mode
  - Measures of Dispersion: Variance, Standard Deviation
```

- Quiz today:

```
  - Mean, Median, Mode
  - Bonus points: Variance, SD
```

- Wednesday: Discussion

```
  - Articles 2 and 3
  - Kahan, D. M., & Corbin, J. C. (2016). A note on the perverse effects of
actively open-minded thinking on climate-change polarization. Research &
Politics, 3(4).
  - Hughes, A. G. (2015). Visualizing inequality: How graphical emphasis
shapes public opinion. Research & Politics, 2(4).
```

Measures of Central Tendency

Measures of central tendency help us:

- reveal patterns
- find the typical measurement
- find the center

Measures of Central Tendency

A few numbers that can summarize the center of measurement

- Mean

- Median
- Mode

Mean

- Symbol: $\bar{x}$
- Not the middle value
- Not the most common
- The center of mass - the sum above equals the sum below
- Formula is $\bar{x} = \frac{\sum X_i}{n}$
- Read that: The mean of X equals the sum of the observations (i) of X divided by the number (n) of observations.

Example 1:

Find the mean of:

1, 7, 3, 4, 5

- $\bar{x} = \frac{\sum X_i}{n}$
- $\bar{x} = \frac{1+7+3+4+5}{5}$
- $\bar{x} = 4$

Example 2:

Find the mean of:

1, 7, 3, 4, 5, 100

- $\bar{x} = \frac{\sum X_i}{n}$
- $\bar{x} = \frac{1+7+3+4+5+100}{6}$
- $\bar{x} = 20$

Median

- Midpoint
- Half observations are greater, half are lower
- Sort the numbers
- Then count
- No formula
- Even observations - midpoint between middle two (mean of the middle two)

Example 1:

Find the median of:

1, 7, 3, 4, 5

- Sort the numbers: 1, 3, 4, 5, 7
- The middle value is 4, so the median is 4.

Example 2:

Find the median of:

1, 7, 3, 4, 5, 100

- Sort the numbers: 1, 3, 4, 5, 7, 100
- The middle two values are 4 and 5, so the median is the mean of these two: $\frac{4+5}{2} = 4.5$.

Mode

- The most common value
- Can be more than one mode (bimodal, multimodal)
- Can be no mode (if all values are unique)
- Not affected by outliers
- The only measure for nominal data
- Just count

Example 1:

Find the mode of:

1, 7, 3, 4, 5

- All values are unique, so there is no mode.

Example 2:

Find the mode of:

1, 7, 3, 4, 5, 7

- The value 7 appears twice, while all other values appear once, so the mode is 7.

Example 3:

Find the mode of:

1, 7, 3, 4, 5, 7, 3

- The values 7 and 3 both appear twice, while all other values appear once, so the modes are 7 and 3 (bimodal).

Measures of Dispersion (Variation or Spread)

- Variance
- Standard Deviation

Spread

- We start with the mean
- Trying to make the picture complete
- How much do the observations vary around the mean?

Potential measure

- Just add up the deviations from the mean: $\sum(X_i - \bar{x})$
- But this always equals zero because the mean is the center of mass

Potential measure 2

- Just add up the absolute value of the deviations from the mean: $\sum|X_i - \bar{x}|$
- Divide this by n to get the average absolute deviation from the mean: $\frac{\sum|X_i - \bar{x}|}{n}$
- This is called the mean absolute deviation (MAD)
- But this is not used much because it is not mathematically tractable
- Not useful for statistical inference such as confidence intervals and hypothesis testing

Potential measure 2: Test question

- Question: There is a much less useful measure of dispersion that is based on the absolute value of deviations from the mean. What is it called?
- Answer: Mean Absolute Deviation, MAD

Variance

- What is the other way we can avoid the problem of deviations from the mean summing to zero?
- Square the deviations from the mean: $\sum (X_i - \bar{x})^2$
- This is called the sum of squared deviations from the mean
- This number is inflated as the number of observations grows…

Variance (Cont.)

- Divide by n to get the average squared deviation from the mean:
- $\frac{\sum (X_i - \bar{x})^2}{n}$
- This is the population variance, σ^2 (sigma squared)
- But we usually don't have measurements for the entire population

Sample Variance

- The population variance is systematically too small because the sample mean is closer to the sample observations than the population mean
- To correct for this bias, we divide by n-1 instead of n to get the sample variance (*Bessel's correction*):
- $\frac{\sum (X_i - \bar{x})^2}{n-1}$

- This is the sample variance, s^2 (s squared)
- This is an *unbiased estimator* of the population variance

Parameters and Statistics

- A *parameter* is a characteristic of a population (e.g., population mean μ, population variance σ^2)
- A *statistic* is a characteristic of a sample (e.g., sample mean $\bar{x}$, sample variance s^2)
- We use statistics to *estimate* parameters

Exam "bonus" question

- Question: We divide by n-1 instead of n to get an unbiased estimator of the population variance. What is this correction called?
- Answer: Bessel's correction

Standard Deviation

- The variance is in squared units, which can be hard to interpret
- To make it easier to work with, we want to get back to the original units
- We take the square root of the variance to get the standard deviation:
- $s = \sqrt{\frac{\sum (X_i - \bar{x})^2}{n-1}}$

or

- $s = \sqrt{s^2}$
- This is the sample standard deviation, s (s)

Summary

- Measures of central tendency: mean, median, mode
- Measures of dispersion: variance, standard deviation
- Variance is the average squared deviation from the mean
- Standard deviation is the square root of the variance
- The sample variance and standard deviation use n-1 in the denominator to correct for bias (Bessel's correction)

Practice Application 1

- If the variance is 100, what is the standard deviation?
- The standard deviation is the square root of the variance, so $s = \sqrt{100} = 10$.

Practice Application 2

- If the standard deviation is 5, what is the variance?
- The variance is the square of the standard deviation, so $s^2 = 5^2 = 25$.

Practice Application 3

- The mean of a sample is 50 and the sum of squared deviations from the mean is 200. If there are 10 observations in the sample, what is the sample variance and standard deviation?
- The sample variance is calculated as $s^2 = \frac{\sum (X_i - \bar{x})^2}{n-1} = \frac{200}{10-1} = \frac{200}{9} \approx 22.22$.
- The sample standard deviation is the square root of the sample variance, so $s = \sqrt{22.22} \approx 4.71$.