POLS 5377 Scope & Method of Political Science

Week 5 Descriptive Statistics

Measures of Dispersion

Healey. (2016) *Statistics: A Tool for Social Research,* Chapter 4

Key Question:

- What is dispersion?
- How to compute and interpret the measures of dispersion: the range (R), the interquartile range (Q), the standard deviation (σ), and the variance (σ^2)?
- How to visualize dispersion via boxplots?

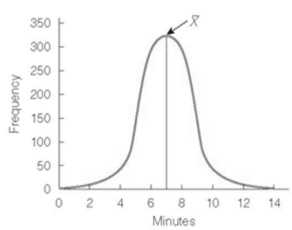
Outline

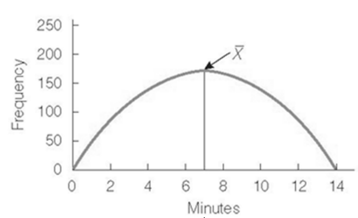
- The Concept of Dispersion
- Range and Interquartile Range
- Standard Deviation and Variance
- Visualizing Dispersion: Boxplots

The Concept of Dispersion

- Two datasets can have same mode, median, and mean, but still very different in terms of the range of dataset.
- Ambulance response time to calls for assistance of two service teams:

A





B

Mean = Median = Mode = 7 minutes

The Concept of Dispersion

- **Dispersion** refers to the variety, diversity, or amount of variation among scores.
- The greater the dispersion of a variable, the greater the range of scores and the greater the differences between scores.

The Concept of Dispersion

- There are four common measures of variability
  - Range (*R*)
  - Interquartile Range (*Q*)
  - Variance (σ^2 for population, S^2 for sample distribution)
  - Standard deviation (σ for population, S for sample distribution)

Range

- Range indicates the distance between the highest and lowest scores in a distribution
- Range (*R*) = High Score – Low Score
- Quick and easy indication of variability
- Can be used with ordinal or interval-ratio variables, but cannot be used with variables measured at the nominal level

25, 25, 30, 30, 35, 35	$R = 35 - 25 = 10$
5, 10, 20, 40, 50, 55	$R = 55 - 5 = 50$

Range

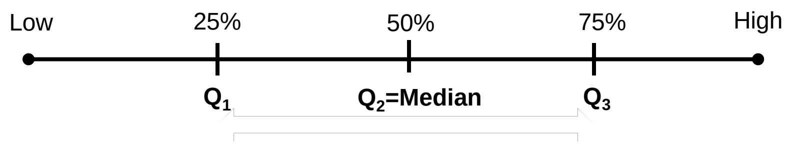
- Limitation:
  - Range only takes the highest and lowest scores into account
  - No information about variation between the high and low scores.

10, 30, 30, 30, 30, 30, 50 → $R = 40$

20, 20, 20, 20, 20, 60, 60 → $R = 40$

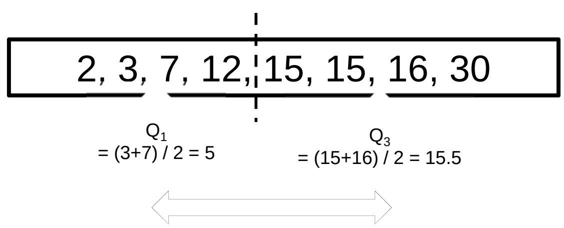
Interquartile Range (*Q*)

- A type of range measure
- Considers only the middle 50% of the cases in a distribution: Q_1 & Q_3 (first quartile and third quartile)
- Avoids some of the problems of the range by only focusing on the middle 50% of scores



$$Q = Q_3 - Q_1$$

Interquartile Range (*Q*)



$$\begin{aligned} Q &= Q_3 - Q_1 \\ &= 15.5 - 5 \\ &= 10.5 \end{aligned}$$

- Limitation: the Interquartile Range is based on only two scores, it fails to yield any information from all of the other scores

Standard Deviation

- The most important and widely used measure of dispersion
- Should be used with interval-ratio variables but is often used with ordinal-level variables
- It is a good measure of dispersion, because:
  1. Use all scores in the distribution
  2. Describe the average or typical deviation of the scores
  3. Increase in value as the distribution of scores becomes more diverse

Standard Deviation & Variance

- The **Variance** and **Standard Deviation** are based on the distance between each score and the mean
- Formulas for variance and standard deviation:

Variance: $\sigma^2 = \frac{\sum(X_i - \bar{X})^2}{N}$

Standard Deviation: $\sigma = \sqrt{\frac{\sum(X_i - \bar{X})^2}{N}}$

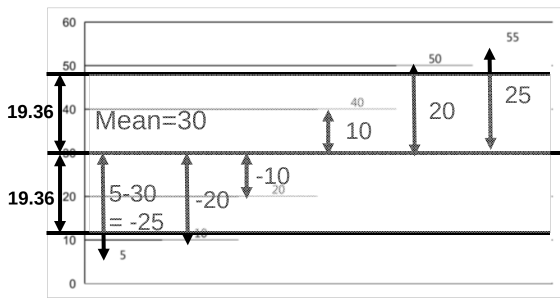
- Standard Deviation: the **square root** of the **Variance**.
- Variance: The average of the **squared differences** from the Mean.

Standard Deviation & Variance

- Example:

5, 10, 20, 40, 50, 55

Mean = (5+10+20+40+50+55)/6 = 30



Variance: σ^2

$$= \frac{(-25)^2+(-20)^2+(-10)^2+10^2+20^2+25^2}{6}$$

$$= \frac{2250}{6} = \mathbf{375}$$

Standard Deviation: $\sigma = \sqrt{375} = \mathbf{19.36}$

Standard Deviation & Variance

- Steps of calculating Variance & Standard Deviation:
  - Calculate the mean ($\bar{X}$)
  - Subtract the mean from each score ($X - \bar{X}$)
  - Square each result $(X - \bar{X})^2$
  - Sum the total $\sum(X - \bar{X})^2$
  - Divide the sum by the total number of cases to get Variance $\frac{\sum(X-\bar{X})^2}{N}$
  - Take the square root of Variance to get the Standard Deviation $\sqrt{\frac{\sum(X-\bar{X})^2}{N}}$

Standard Deviation & Variance

(A) 25, 25, 30, 30, 35, 35

X_i	$(X-\bar{X})$	$(X-\bar{X})^2$
25	-5	25
25	-5	25
30	0	0
30	0	0
35	5	25
35	5	25
$\bar{X}=30$		$\sum(X-\bar{X})^2$=100
$\sigma^2=\frac{\sum(X-\bar{X})^2}{N}=16.67$		$\sigma=\sqrt{16.67}$ =4.08

5, 10, 20, 40, 50, 55 (B)

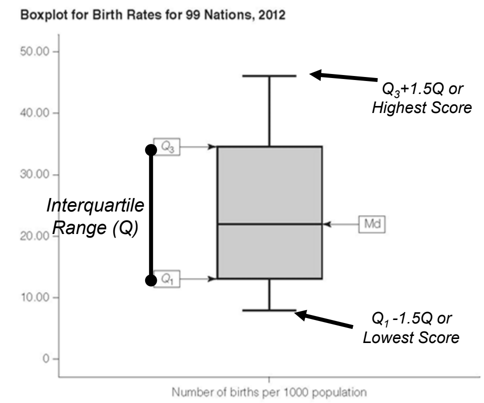
X_i	$(X-\bar{X})$	$(X-\bar{X})^2$
5	-25	625
10	-20	400
20	-10	100
40	10	100
50	20	400
55	25	625
$\bar{X}=30$		$\sum(X-\bar{X})^2$=2250
$\sigma^2=\frac{\sum(X-\bar{X})^2}{N}=375$		$\sigma=\sqrt{375}=19.36$

Standard Deviation & Variance

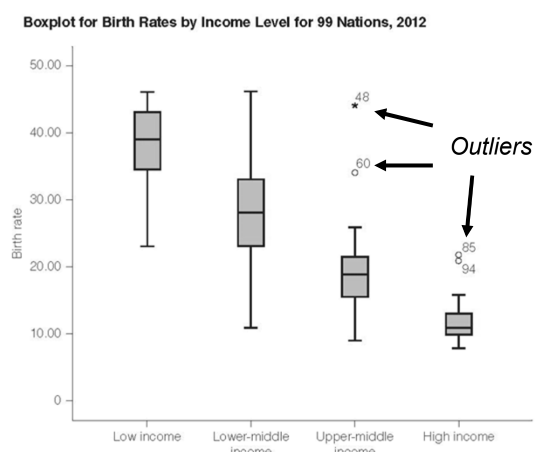
- Report:
  - In Team A, the mean work load was 30 cases a week with a standard deviation of 4.08.
  - In Team B, the mean work load was 30 cases a week with a standard deviation of 19.36.
- The size of the standard deviation is meaningful only in comparison with the mean.
  - For example, if the mean is 1000, a standard deviation of 10 is small; if the mean is 15, a standard deviation of 10 is large.
- The larger the standard deviation in relation to mean, the more dispersed are the values of the distribution.

Visualizing Dispersion: Boxplots

- Boxplots (or box and whiskers plots) provide helpful ways to visualize and analyze dispersion.



Visualizing Dispersion: Boxplots



After this lecture:

You should be able to:

- Explain the concept of dispersion
- Compute measures of dispersion: Range (*R*), Interquartile Range (*Q*)
- Compute and interpret Standard Deviation(σ) and Variance(σ^2)
- Interpret and Draw boxplots