DISPERSION & Z SCORES

Dr. Derek Borman

Understanding variability and spread

INTRODUCTION TO STATISTICS

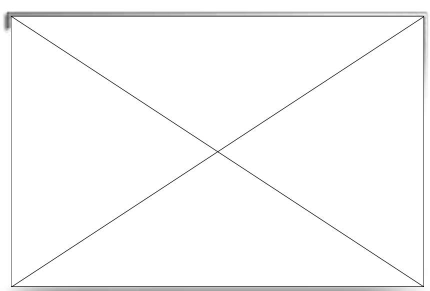
KEY OBJECTIVES

- Be able to define and compute the measures of dispersion covered in the chapter–the range, average deviation, variance, and standard deviation.
- Be able to define, computer and interpret z scores or standard scores.
- Understand the relationship between standard deviation and z scores.

Why is it important to understand dispersion within a single distribution?

- What is the "center" of the data?
- Suppose that we wanted to open a shop that sells shoes?
  - What are the mean, median and mode for sizes of shoes out there?
  - Once we have these measures of central tendency, can we go out and start buying and selling shoes?

Why is it important to understand dispersion when comparing distributions?

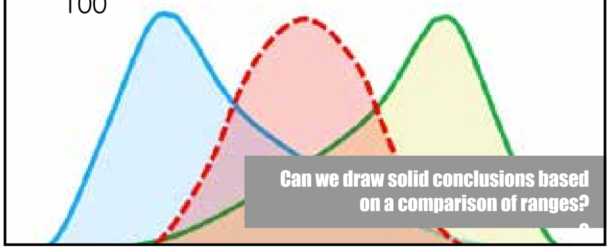


Range

- The difference between largest and smallest data point (HS – LS).
- Quick way to find spread.
- Highly affected by outliers (i.e., extreme scores).
- Best for symmetric data with no outliers.

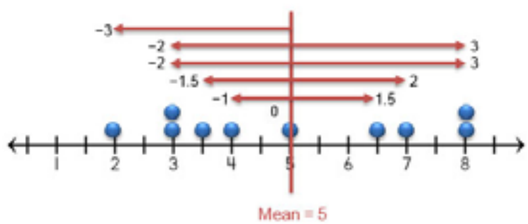
Two Stats classes take the same test. What is the Range for both classes?

- Class A: 13, 23, 33, 43, 53, 63, 73, 83, 93, 100
- Class B: 13, 85, 85, 86, 87, 87, 88, 88, 89, 100



Can we draw solid conclusions based on a comparison of ranges?

Average Deviation

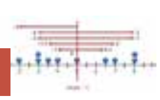


Average Deviation

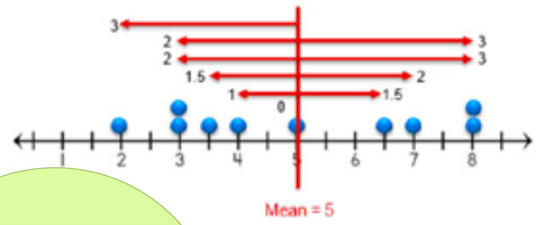
- Sum distances between each raw score and mean and divide N

$$\frac{\sum (X - \overline{X})}{N}$$

So, what's the problem with average deviation?



ABSOLUTE AVERAGE DEVIATION



$\sum |X-\bar{X}|$

So, what's the problem with absolute average deviation?

Variance

- How do we avoid 0?
  - Find variance: average deviation of squared differences from the mean
    - The numerator is referred to as the "sum of squares" or SS.

$$s^2 = \frac{\sum (X-\bar{X})^2}{N-1}$$

Variance

At a recent fraternity hazing party, fraternity pledges were told that whoever could eat the most goldfish would be admitted into the fraternity. The number of goldfish eaten by each hopeful is listed below. Find the variance?

25, 24, 23, 22, 20, 18, 17, 16, 14, 10, 8

$$s^2 = \frac{\sum (X-\bar{X})^2}{N-1}$$

Now, Try it With the Computational Formula

$$s^2 = \frac{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}{N-1}$$

Variance

- Highly affected by outliers. Best for symmetric data.
- If measuring variance of population, denoted by σ^2 ("sigma-squared").
- If measuring variance of sample, denoted by s^2 ("s-squared").
- Problem is units are squared.

Standard deviation or s

- Sample standard deviation (s) is square root of variance
- Units are original units
- Average distance of scores from mean
- s = approximately R/4
- Large s indicates error (invalid, unreliable data)

$$s = \sqrt{\frac{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}{N-1}} \qquad s = \sqrt{\frac{\Sigma(X-\bar{X})^2}{N-1}}$$

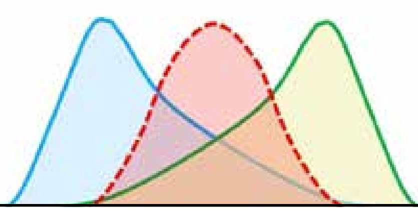
$$s = \sqrt{\frac{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}{N - 1}}$$

WHY N-1 in the denominator?

- Why not N? What does reducing the denominator do to our standard deviation?
  - Hint: Visualize the standard deviation curve.
- STATISTICS ARE ESTIMATES!

Choosing Appropriate Measure of Dispersion

- If data symmetric, with no serious outliers, use range and standard deviation
  - Skewed data described using other statistics



Statistics: The only class where you can get A's…for catching some Z's.

Z Scores: How We Standardize Data

- Z indicates how a specific score deviates from the mean, in s units
- S provides general index of deviation
- Positive and negative z scores
- Z (standardized scores) allow us to compare individuals across distributions

How do we calculate z?

- Suppose a mean of 10 and a standard deviation of 2. How many standard deviation units (z-score units) away from the mean is a raw score of 14? How about 9? 5?

$$z = \frac{X - \bar{X}}{s}$$

Can we compare z scores?

Year	HR Mean	Standard Deviation	Name	Player HR	Z-Score
1920	4.85	7.27	Babe Ruth	54	
1922	6.31	7.18	Rogers Hornsby	42	
1932	8.59	10.32	Jimmie Foxx	58	
1990	11.12	8.74	Cecil Fielder	51	
1949	10.87	9.78	Ralph Kiner	54	
1998	15.44	12.48	Mark McGwire	70	
2001	18.03	13.37	Barry Bonds	73	

Can we compare z scores?

Year	HR Mean	Standard Deviation	Name	Player HR	Z-Score
1920	4.85	7.27	Babe Ruth	54	6.76
1922	6.31	7.18	Rogers Hornsby	42	4.97
1932	8.59	10.32	Jimmie Foxx	58	4.79
1990	11.12	8.74	Cecil Fielder	51	4.56
1949	10.87	9.78	Ralph Kiner	54	4.41
1998	15.44	12.48	Mark McGwire	70	4.37
2001	18.03	13.37	Barry Bonds	73	4.11

- What if we have a z score but need to figure out the raw score?
- We use the following formula:

$$X = zs + \bar{X}$$

Suppose WNBA average height is 72 inches, standard deviation is 2.3. What is the height of a player who is 2.6 standard deviation units above the mean? How about the height of a player with a z-score of -1.47?