Statistics Notations Population vs Sample

QL 1.1

Learning Outcomes

By the end of today, you should be able to

1. Compare and contrast sample vs population variance and standard deviations

2. Identify the position of a data value in a data set, using z-score

Population vs Sample

- A **population** includes **all** of the elements from a set of data.

- A **sample** consists one or more observations drawn from the population.

- The population size is shown by **N**.

- The sample size is shown by of **n**.

- The mean of a population is denoted by the symbol μ but the mean of a sample is denoted by the symbol $\overline{X}$

Population Mean	Sample Mean
$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$	$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$
N = number of items in the population	n = number of items in the sample

Population vs Sample Variance

Population Variance	Sample Variance
$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}$$	$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$
σ^2 = population variance	s^2 = sample variance
x_i = value of ith element	x_i = value of ith element
μ = population mean	$\bar{x}$ = sample mean
N = population size	n = sample size

Population vs Sample variance

5 min

a. A population consists of four observations: {1, 3, 5, 7}. What is the variance?

b. Now assume simple random sample consists of four observations: {1, 3, 5, 7}. Based on these sample observations, what is the best estimate of the variance of the population?

Answer

Measure of Position: z-score
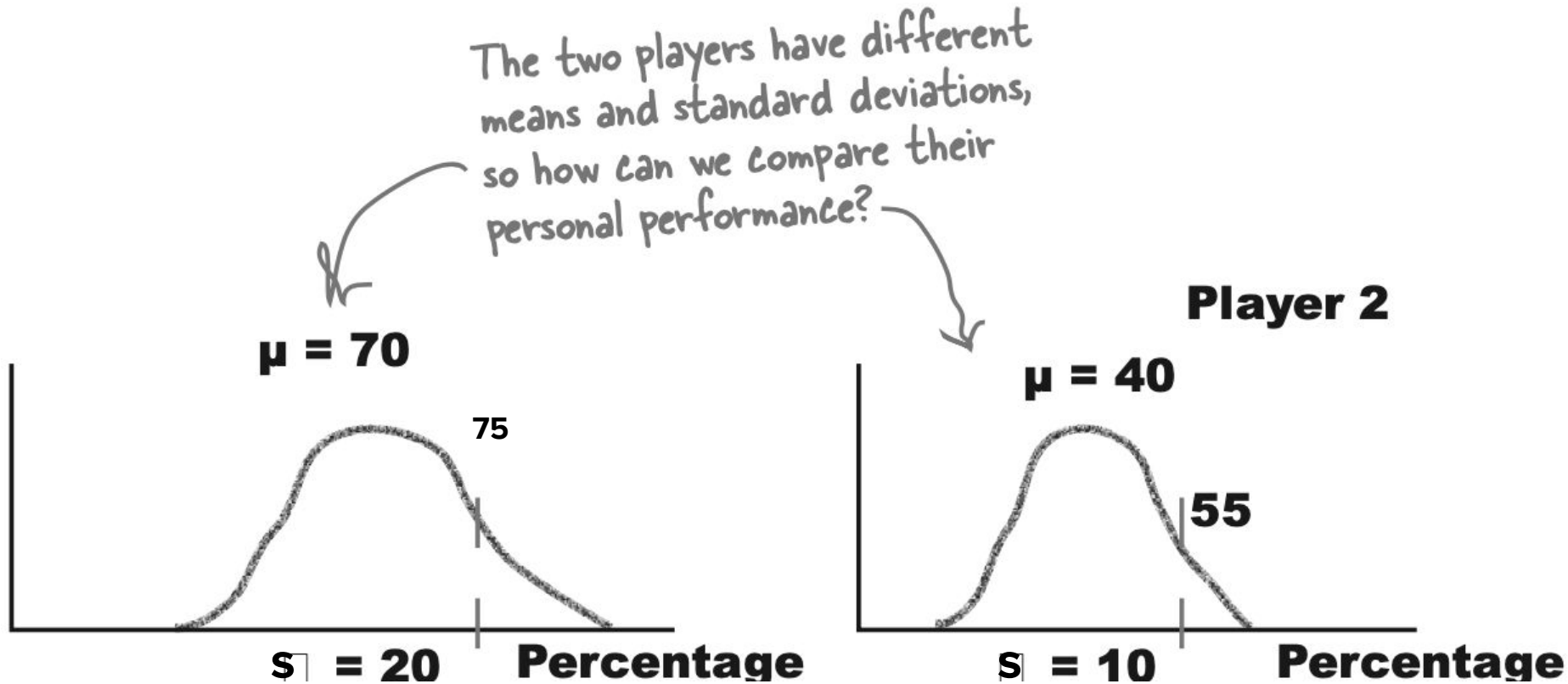
- We have

  - measures of **central tendency**

    - Mean, median, mode

  - measures of **variation**

    - Range, variance, standard variation

  - measures **position (=location)**

    - Percentiles and **Standard scores (z scores)**

Imagine a situation in which you have two basketball players of different ability. The first player gets the ball into the net an average of 70% of the time, and he has a standard deviation of 20%. The second player has a mean of 40% and a standard deviation of 10%.

In a particular practice session, Player 1 gets the ball into the net 75% of the time, and Player 2 makes a basket 55% of the time. Which player does best against their personal track record?

The two players have different means and standard deviations, so how can we compare their personal performance?

Player 2

$\mu = 70$

75

$s_1 = 20$ **Percentage**

$\mu = 40$

55

$s_1 = 10$ **Percentage**

Standard Score (z-score)

A standard score or z score tells how many standard deviations a data value is above or below the mean. A z score or standard score for a value is obtained by
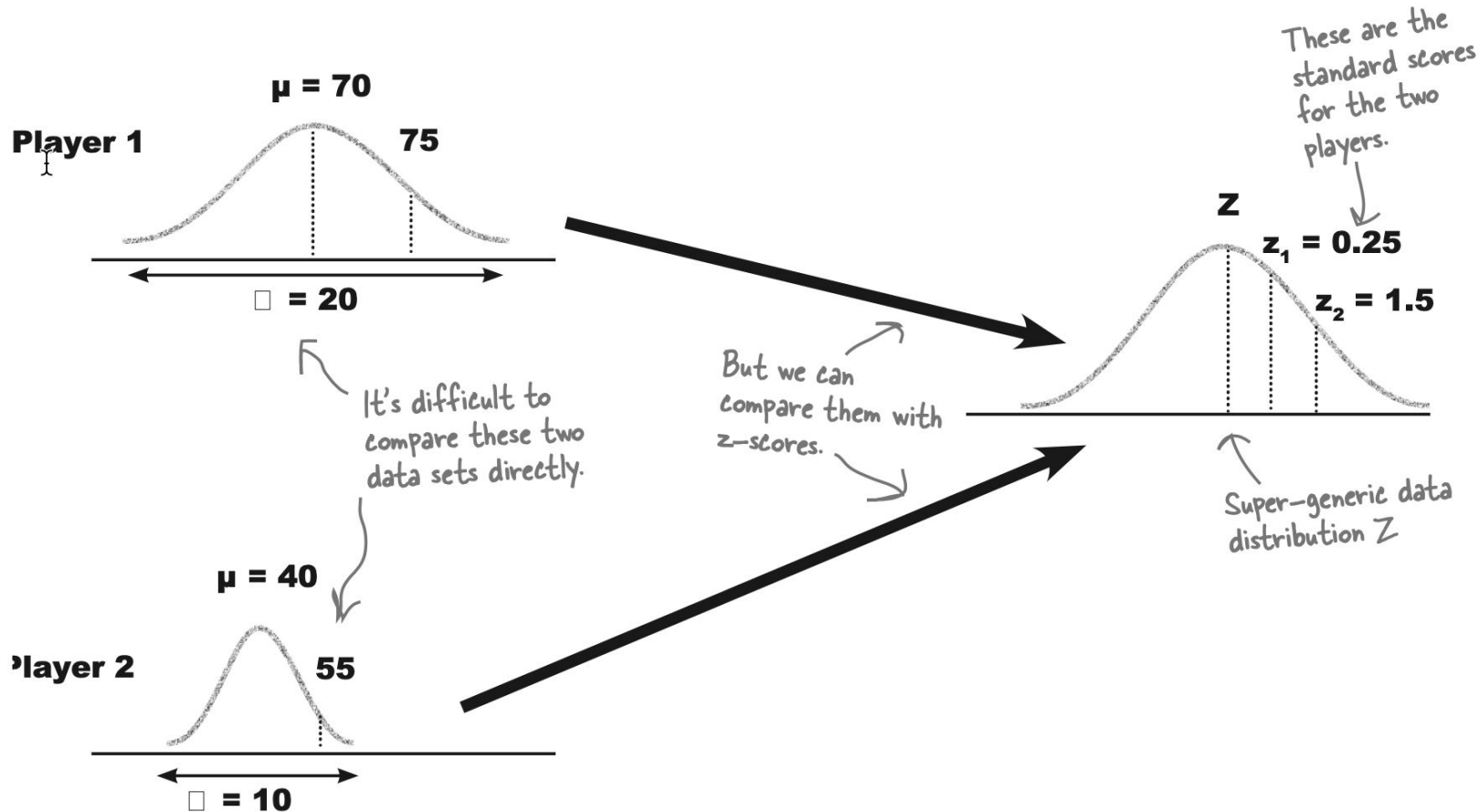
$$z = \frac{value - mean}{standard\ \ deviation}$$

For populations, the formula is

$$z = \frac{X - \mu}{\sigma}$$

For samples, the formula is:

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

Let's

Player 1

$\mu = 70$

75

$\sigma = 20$

It's difficult to compare these two data sets directly.

But we can compare them with z-scores.

$\mu = 40$

Player 2

55

$\sigma = 10$

These are the standard scores for the two players.

Z

$z_1 = 0.25$

$z_2 = 1.5$

Super-generic data distribution Z

Standard Scores (z-Scores)

- A student scored 65 on a calculus test that had a mean of 50 and a standard deviation of 10. She scored 30 on a history test with a mean of 25 and a standard deviation of 5. Compare her relative positions to other students on the two tests.

Solution

First, find the z scores. For calculus the z score is

$$z = \frac{X - \overline{X}}{s} = \frac{65 - 50}{10} = 1.5$$

For history the z score is

$$z = \frac{30 - 25}{5} = 1.0$$

14

z score

5 mins

Find the z score for each test, and state which is higher.

Test A	$X = 38$	$\overline{X} = 40$	$s = 5$
Test B	$X = 94$	$\overline{X} = 100$	$s = 10$

[Answer](#)