## 3.5: Descriptive Measures for Populations; Use of Samples
## (also *z*-scores and measures of position)

**Population mean, population variance, and population standard deviation:**

Recall:

Population mean:

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

Sample mean:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

Population variance:

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n}$$

Sample variance:

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}$$

Population standard deviation:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n}}$$

Sample standard deviation:

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}}$$

**Parameters and statistics:**

Definition:

A *parameter* is a descriptive measure for a population.
A *statistic* is a descriptive measure for a sample.

**The z-score:**

The z-score for a data point represents the number of standard deviations that lie between the data point and the mean. The z-score is sometimes known as the standardized value, and it allows us to compare data points from different distributions.

You can think of the z-score as representing the "distance from the mean," with distance measured in standard deviations. A positive z-score indicates the data point lies above the mean; a negative z-score indicates the data point lies below the mean.

Therefore, for bell-shaped distributions (from Empirical Rule),

- about 68% of the data points have z-scores between –1 and 1;
- about 95.7% of the data points have z-scores between –2 and 2;
- about 99% of the data points have z-scores between –3 and 3.

For all distributions (from Chebyshev's Rule),

- at least 75% of the data points have z-scores between –2 and 2;
- at least 88.9% of the data points have z-scores between –3 and 3.

<u>The z-score (or standardized score)</u>:

The z-score for a value $x$ is

$$z = \frac{x-\mu}{\sigma} \text{ (for a population), or}$$

$$z = \frac{x-\bar{x}}{s} \text{ (for a sample), where}$$

$\mu$ and $\sigma$ are the population mean and standard deviation, or
$\bar{x}$ and $s$ are the sample mean and standard deviation.

<u>Note</u>: The z-score is unitless. All distributions of z-scores have mean 0 and standard deviation 1.

**<u>Example 1:</u>** Suppose a data set has mean 52 and standard deviation 8. Find the z-scores for the scores 44, 64, 38, and 52.

Assume it's a sample: $\bar{x} = 52$
$s = 8$

see next page

z-score for 44: $z = \frac{x-\bar{x}}{s} = \frac{44-52}{8} = \frac{-8}{8} = -1$

44 is 1 std. dev. below the mean.

z-score for 64: $z = \frac{x-\bar{x}}{s} = \frac{64-52}{8} = \frac{12}{8} = 1.5$

(so 64 is $1\frac{1}{2}$ std devs above the mean.)

For $x = 38$, $z = \frac{38-52}{8} = \frac{-14}{8} = -\frac{7}{4} = \boxed{-1.75}$

3.5.3
3.4.3

For $x = 52$, $z = \frac{52-52}{8} = \frac{0}{8} = \boxed{0}$

**Example 2:** In 2014, the mean of the ACT mathematics test was 20.9 and the standard deviation was 5.3. In the same year, the mean of the SAT mathematics test was 513 and the standard deviation was 120. Suppose Tamara, a high school student, received a score of 24 on the ACT mathematics test, and 600 on the SAT mathematics test. On which test did she perform better?

(ACT data from the National Center for Education Statistics, https://nces.ed.gov/programs/digest/d14/tables/dt14_226.50.asp?current=yes; SAT data from the College Board, https://www.collegeboard.org/program-results/2014/sat)

We calculate her z-score on each test:

ACT her score: $x = 24$
$\mu = 20.9$
$\sigma = 5.3$

$$z = \frac{x-\mu}{\sigma} = \frac{24-20.9}{5.3} \approx 0.585$$

SAT her score: $x = 600$
$\mu = 513$
$\sigma = 120$

$$z = \frac{x-\mu}{\sigma} = \frac{600-513}{120} = 0.725$$

Relative to the national averages, she scored better on the SAT (because her z-score is higher)