

Section 3.4

Measures of Position and Outliers

Objectives

By the end of this lesson, you will be able to...

1. determine and interpret z-scores
2. determine and interpret percentiles
3. determine and interpret quartiles
4. check a set of data for outliers

In Sections 3.1 and 3.2, we discussed ways to describe a "typical" individual in a population or sample. In this next section, we'll talk about ways to describe an *individual* in relation to the population.

z-Scores

Example 1

It's fairly common for upper-level statistics courses to have both undergraduate and graduate students. Given the exam scores listed below, can you determine which score is better relative to its peers, the undergraduate score of 83 or the graduate score of 88?

undergraduate student scores	graduate student scores
65 89 84 75	82 90 95 72
52 78 92 80	78 88 92 89
76 72 83 79	

Actually, to answer this question, we need more information. In particular, we need a new way to describe *relative position*.

The **z-score** represents the number of standard deviations a data value is from the mean.

Population z-Score Sample z-Score

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

$$z \approx \frac{x - \bar{x}}{s}$$

$$z = \frac{83 - 77.1}{10.73} \approx 0.55$$

With a sample mean of 85.75 and a standard deviation of 7.78, the graduate has a z-score of:

$$z = \frac{88 - 85.75}{7.76} \approx 0.29$$

Since the undergraduate is more than 1/2 of a standard deviation above the mean ($z = 0.55$), that's a better relative score.

Note: You may have noticed that I went to the hundredths place for these z-scores. That's standard practice.

Key: We use z-scores when we want to compare two individuals from different populations, *relative* to their respective populations.

Percentiles

If you've ever taken a standardized exam like the [PSAT](#), [SAT](#), or [ACT](#), you've seen in the report something about your **percentile**.

The **k th percentile**, denoted P_k , of a set of data divides the lower $k\%$ of a data set from the upper $(100-k)\%$.

Percentile ranks are used in a variety of fields:

- Special Education - students scoring below a certain percentile on specific tests qualify for services.
- [Physicians](#) - doctors usually track a child's weight and height and compare the growth to that of other children of the same age.

Unfortunately, there's no universally accepted way to calculate percentiles. Most software packages and calculators use a method similar to the one below (from your text), but you should be aware of the possibility of others.

Determining the k th percentile, denoted P_k

Step 1: Arrange the data in ascending order.

Step 2: Compute an index i using the formula

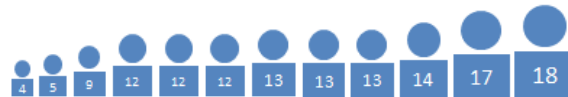
$$i = \left(\frac{k}{100}\right)(n + 1)$$

Step 3:

- If i is an integer, the k th percentile, denoted P_k , is the i th value.
- If i is not an integer, the k th percentile is the mean of the observations on either side of i .

Example 2

Let's go back to the Jackson cousins we saw in [Example 2](#) in Section 3.1. Suppose this time we add all the cousins, from little Zander at age 4 to Mae, who at age 18 is entering her first year at college.

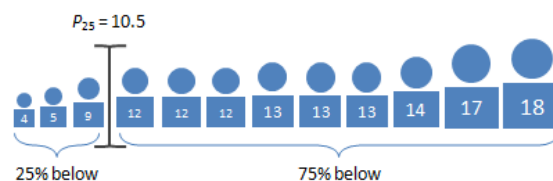


Use the strategy above to find the 25th percentile by age.

Using the formula above, we calculate i as:

$$i = \left(\frac{25}{100}\right)(12 + 1) = \left(\frac{1}{4}\right)(13) = 3.25$$

Since i is not an integer, we average the 3rd and fourth values to get 10.5 yrs.



Technology

Here's a quick overview of the formulas for finding percentiles in StatCrunch.

1. Select **Stat > Summary Stat > Columns**.
2. Select the variable you want to summarize (e.g., "Heights")--leave everything else as is for now.
3. Click "Next".
4. Deselect any statistics that you do not want calculated
5. Enter the percentile you wish to calculate in the "Percentile" box.
6. Click "Calculate" and another window with these numbers calculated will pop up.

Note: Some software like Microsoft Excel interpolates instead of taking a simple average when calculating percentiles, so the results may differ slightly.

Determining the Percentile of a Data Value

The last thing we need to do with percentiles is to figure out the percentile of a particular individual. For example, if your Composite ACT score is a 28, what percentile does that leave you?

As before, there is no universally accepted way to calculate percentiles, but the following (from your text) is very common.

Finding the Percentile that Corresponds to a Data Value

Step 1: Arrange the data in ascending order.

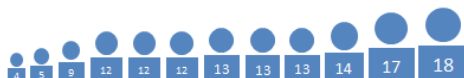
Step 2: Use the following formula to find the percentile of the value, x .

$$\text{percentile of } x = \frac{\text{number of data values less than } x}{n} * 100$$

Round this number to the nearest integer.

Example 3

Consider again the Jackson cousins we looked at in Example 2 above.



What is the percentile rank of James, the 14-year-old?

[reveal answer]

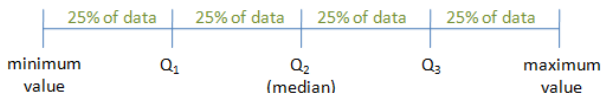
Using the formula above, we calculate i as:

$$\text{percentile of James} = \frac{9}{12} * 100 = 75$$

So James is the 75th percentile.

Quartiles

As the name implies, **quartiles** divide the data into four equal parts. Therefore the first quartile, Q_1 , is the 25th percentile, the second quartile, Q_2 is the 50th percentile (or the median), and the third quartile, Q_3 , is the 75th percentile.



Example 4

Let's consider one of the sets of hypothetical exam scores we looked at in Section 3.2.

48	57	58	65	68	69	71	73	73
74	75	77	78	78	78	79	80	85
87	88	89	89	89	95	96	97	99

Find the quartiles.

[reveal answer]

$$Q_1: i = \left(\frac{25}{100}\right)(27 + 1) = \left(\frac{1}{4}\right)(28) = 7$$

So the 1st quartile is x_7 , or **71**.

Q_2 : Since there are 27 observations, the median is x_{14} , or 78.

$$Q_3: i = \left(\frac{75}{100}\right)(27 + 1) = \left(\frac{3}{4}\right)(28) = 21$$

So the 3rd quartile is x_{21} , or **89**.

Technology

Here's a quick overview of the formulas for finding quartiles in StatCrunch.

1. Select **Stat > Summary Stat > Columns**.
2. Select the variable you want to summarize (e.g., "Heights")--leave everything else as is for now.
3. Click "Next".
4. Deselect any statistics that you do not want calculated
5. Click "Calculate" and another window with these numbers calculated will pop up.

Note: Some software like Microsoft Excel interpolates instead of taking a simple average when calculating percentiles, so the results may differ slightly.

Checking for Outliers

One good use of quartiles is they give us a sense of what values might be extreme. In Statistics, we call these values **outliers**. There are various ways to check for outliers. Most depend on the distribution and often can only characterize observations as *possible* outliers. A common technique used is the following:

Checking for Outliers by Using Quartiles

Step 1: Determine the first and third quartiles

Step 2: Compute the inter-quartile range: $IQR = Q_3 - Q_1$

Step 3: Determine the fences.

$$\text{Lower fence} = Q_1 - 1.5(IQR)$$

$$\text{Upper fence} = Q_3 + 1.5(IQR)$$

Step 4: If a value is less than the lower fence or greater than the upper fence it is considered an outlier.

Example 5

Let's look at those same exam scores we used in Example 4.

48	57	58	65	68	69	71	73	73
74	75	77	78	78	78	79	80	85
87	88	89	89	89	95	96	97	99

Use the above method to determine if there are any outliers.

[\[reveal answer \]](#)

$$IQR = Q_3 - Q_1 = 89 - 71 = 18$$

$$\text{Lower fence} = Q_1 - 1.5(IQR) = 71 - 1.5(18) = 44$$

$$\text{Upper fence} = Q_3 + 1.5(IQR) = 89 + 1.5(18) = 116$$

Since no values lie outside the fences, there are no outliers.