Chapter 3: Data Description

Diana Pell

Section 3.3: Measures of Position

A percentile, or percentile rank, of a data value indicates the percent of data values in a set that are below that particular value.

If your percentile rank on SAT was 60%, that means that 60% of all students who took the SAT scored lower than you did.

Percentile Formula

The percentile corresponding to a given value \( X \) is computed by using the following formula:

\[
\text{Percentile} = \frac{\text{(number of values below } X) + 0.5}{\text{total number of values}} \times 100
\]

Exercise 1. Suppose you score 77 on a test in a class of 10 people, with the 10 scores listed below. What was your percentile rank?

\[
93 \ 82 \ 64 \ 75 \ 98 \ 52 \ 77 \ 88 \ 90 \ 71
\]

\[
\text{Percentile} = \frac{4 + 0.5}{10} \times 100 = 45\text{th percentile}
\]
Exercise 2. The weights in pounds for the 12 members of a college gymnastics team are below. Find the percentile rank of the gymnast who weighs 97 pounds.

\[
\frac{\frac{7 + 0.5}{12} \times 100}{12} = 62.5 \approx 63\text{rd percentile}
\]

Finding a Data Value Corresponding to a Given Percentile

1. Arrange the data in order from lowest to highest.

2. Substitute into the formula

\[
c = \frac{n \cdot p}{100}
\]

where \( n \) = total number of values and \( p \) = percentile

3. If \( c \) is not a whole number, round up to the next whole number. Starting at the lowest value, count over to the number that corresponds to the rounded-up value. If \( c \) is a whole number, use the value halfway between the \( c \text{th} \) and \((c + 1)\text{st}\) values when counting up from the lowest value.

Exercise 3. Find the value corresponding to the 25th percentile.

18, 15, 12, 6, 8, 2, 3, 5, 20, 10

\[
c = \frac{10 \cdot 25}{100} = 2.5 \approx 3
\]

(not a whole # \rightarrow round up)

\[
2 \quad 3 \quad 5 \quad 6 \quad 8 \quad 10 \quad 12 \quad 15 \quad 18 \quad 20
\]

\[
5 \text{ corresponds to 25th percentile}
\]
Exercise 4. Find the value that corresponds to the 60th percentile.

\[ 18, 15, 12, 6, 8, 2, 3, 5, 20, 10 \]

\[ 2, 3, 5, 6, 8, 10, 12, 15, 18, 20 \]

\[ c = \frac{n \cdot p}{100} = \frac{10 \cdot 60}{100} = 6 \rightarrow \text{whole # \& use value halfway w/ 6th and 7th values.} \]

Another statistical measure we will study is the **quartile**, which divides a data set into quarters. The **second quartile** is the same as the median, and divides a data set into an upper half and a lower half. The **first quartile** is the median of the lower half, and the **third quartile** is the median of the upper half. We use the symbols \( Q_1, Q_2, \) and \( Q_3 \) for the first, second, and third quartiles respectively.

The **interquartile range (IQR)** is the difference between the third and first quartiles.

\[ IQR = Q_3 - Q_1 \]

**The Five-Number Summary and Boxplots**

A boxplot can be used to graphically represent the data set. These plots involve five specific values:

1. The lowest value of the data set (i.e., minimum)
2. \( Q_1 \)
3. The median (\( Q_2 \))
4. \( Q_3 \)
5. The highest value of the data set (i.e., maximum)

These values are called a **five-number summary** of the data set.

A boxplot is a convenient way of graphically depicting groups of numerical data through their quartiles.
Exercise 5. The number of meteorites found in 10 states of the United States is

\[89, 47, 164, 296, 36, 215, 138, 78, 48, 39\]

a) Find the five-number summary.

\[
\begin{align*}
\text{min} &= 30 \\
\text{max} &= 296 \\
\overline{Q_2} &= \frac{78 + 84}{2} = 81.5 \\
Q_1 &= 47 \\
Q_3 &= 164
\end{align*}
\]

b) Construct a boxplot for the data.

![Boxplot](image)

c) Find the IQR.

\[IQR = Q_3 - Q_1 = 164 - 47 = 117\]

d) What does the position of the box tell you about the data set?

The distribution is skewed to the right.

An outlier is an extremely high or an extremely low data value when compared with the rest of the data values.

Procedure for Identifying Outliers

1. Find the interquartile range: \(IQR = \overline{Q_3} - Q_1\)

2. Multiply the IQR by 1.5
3. Subtract the value obtained in step 2 from $Q_1$ and add the value obtained in step 2 to $Q_3$.

4. Check the data set for any data values that fall outside the interval.

**Exercise 6.** Check the following data set for outliers.

5, 6, 12, 13, 15, 18, 22, 50

\[
IQR = Q_3 - Q_1 = 20 - 9 = 11
\]

\[1.5 \times (IQR) = 16.5\]

\[9 - 16.5 = -7.5\]

\[9 + 16.5 = 36.5\]

Data values that fall outside of the interval from -7.5 to 36.5 are considered outliers.

- 50 is an outlier

* Nonresistant statistics (affected by outliers): mean, std. deviation

* Resistant statistics (less affected by outliers): median, IQR