3.3 - Measures of Position

Z-Scores

Who is taller, a man 73 inches tall or a woman 68 inches tall? The obvious answer is that the man is taller. However, men are taller than women on the average. Let’s ask the question this way: Who is taller relative to their gender, a man 73 inches tall or a woman 68 inches tall?

The **z-score** of an individual data value tells how many _________________ that value is from its population mean.

Let $x$ be a value from a population with mean $\mu$ and standard deviation $\sigma$. The z-score for $x$ is

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

Practice

1. A National Center for Health Statistics study states that the mean height for adult men in the U.S. is $\mu = 69.4$ inches, with a standard deviation of $\sigma = 3.1$ inches. The mean height for adult women is $\mu = 63.8$ inches, with a standard deviation of $\sigma = 2.8$ inches. Who is taller relative to their gender, a man 73 inches tall, or a woman 68 inches tall?

2. Eric proudly tells his brother Bruce that he got 94 points on his last math exam, which had an average of 73 points and a standard deviation of 9. Bruce says that he did even better on his math exam, on which he got 96 points, and this exam had an average of 79 points and a standard deviation of 7. Who did better on his exam relative to their class scores?

3. Suppose Eric's classmate got a z-score of -1.7 on his math exam. What was his exam score?
Percentiles

Percentiles denoted $P_1, P_2, \ldots, P_{99}$, divide sorted data into \textbf{99 equal parts}.

\textit{Draw a picture:}

The $p^{\text{th}}$ percentile is the value in a data set that has about $p\%$ of the data values smaller than $p$ and $(100-p)\%$ values that are greater than $p$.

\textit{ex.} When my daughter Linnéa was born, the doctor told me her length was in the 98.8th percentile for newborn girls. What does that mean?

To find the approximate value of the $p^{\text{th}}$ percentile:

1. sort the data in increasing order
2. calculate $L = \left( \frac{p}{100} \cdot n \right)$
3. If $L$ is a whole number, the $p^{\text{th}}$ percentile is the average of the number in position $L$ and the number in position $L+1$.
   If $L$ is NOT a whole number, the $p^{\text{th}}$ percentile is the number in the position of the next whole number higher than $L$.

\begin{align*}
\text{Percentile rank of } x &= \left( \frac{\text{Number of values less than } x}{n} + 0.5 \right) \cdot 100 \\
\text{(where } x \text{ is some number in the data set)}
\end{align*}

Round the result to the nearest whole number.

\textit{ex.} Given the following data set

\[\begin{array}{cccccccccc}
15 & 9 & 12 & 11 & 7 & 6 & 9 & 10 & 14 & 3 & 6 & 5 \\
\end{array}\]

(a) Calculate the approximate value of the 55th percentile.

(b) Find the percentile rank of 7.
Quartiles

Quartiles denoted $Q_1$, $Q_2$, and $Q_3$, divide sorted data into __________ equal parts.

<table>
<thead>
<tr>
<th>25%</th>
<th>25%</th>
<th>25%</th>
<th>25%</th>
</tr>
</thead>
<tbody>
<tr>
<td>(minimum)</td>
<td>$Q_1$</td>
<td>$Q_2$</td>
<td>$Q_3$ (maximum)</td>
</tr>
<tr>
<td>(median)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$Q_2$ is the 50th percentile (median).
$Q_1$ is the 25th percentile.
$Q_3$ is the 75th percentile.

The five-number summary of a data set consists of: Minimum value, $Q_1$, $Q_2$, $Q_3$, and Maximum value.

$IQR = \text{Interquartile Range} = Q_3 - Q_1$

The IQR method allows us to determine which values are outliers. Outliers are data values that are below $Q_1 - 1.5 \cdot IQR$, the lower outlier boundary, or above $Q_3 + 1.5 \cdot IQR$, the upper outlier boundary.

Find the five number summary and the IQR for the given data sets below. Determine if they have any outliers.

ex. 1 3 6 7 9 9 10 14 15

ex. 3 2 10 8 2 9

Calculator use:
Boxplots

A boxplot is a graphic presentation of data using the five number summary, the IQR, and outliers.

Lower outlier boundary = \( Q_1 - 1.5 \times IQR \)
Upper outlier boundary = \( Q_3 + 1.5 \times IQR \)

How to draw a box-and-whisker plot:

- Draw a number line, such that all numbers in the data set are covered.
- Draw a box above the number line, such that its left side is at \( Q_1 \) and the right side at \( Q_3 \).
  - Draw a vertical line at \( Q_2 \) also.
- Draw whiskers (horizontal lines) to join the box and the smallest and largest value resp. within the two outlier boundaries.
- Plot any outliers (values outside of outlier boundaries).

ex. The time (in minutes) that a student spent in the laundromat in a week, for 15 randomly selected weeks, is as follows:

\[
\begin{array}{cccccccc}
72 & 62 & 84 & 73 & 107 & 81 & 93 & 72 \\
135 & 77 & 85 & 67 & 90 & 83 & 112 \\
\end{array}
\]

(a) Prepare a box-and-whisker plot.

(b) Is the data skewed?

(c) Does the data contain any outlier?
A boxplot can help us better see the distribution of the data, such as the center, spread, skewness, and outliers.

(a) Normal (bell-shaped) distribution
1000 heights (in.) of women

(c) Skewed distribution
Incomes (thousands of dollars) of 1000 statistics professors

We can also use boxplots to visually compare two or more data sets by placing them right above each other.