Measures of Position

Summary

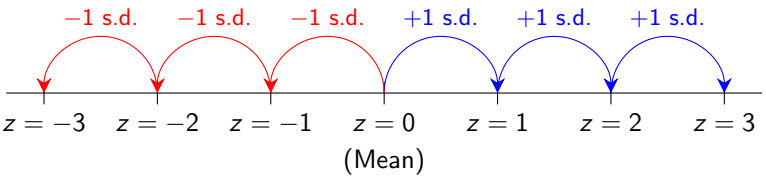
1. A z-score tells you how many standard deviations above or below the mean a data value is.
2. Percentiles tell what percent of values are below a given value.
3. Box plots and five-number summaries can be used to detect outliers in data sets.

Z-Scores

z-score

A **z-score** (a.k.a. **standard score**) measures how many standard deviations a data value, x, is from the mean of the data set.

- A positive z-score $\rightarrow$ above average value.
- A negative z-score $\rightarrow$ below average value.
- A z-score of 0 $\rightarrow$ exact average value.



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

- Can compare two different data sets that use different scales of measurement.
- "Usual" data values have z-scores between -2 and 2.

Example 1. The mean SAT score is 1059 with a standard deviation of 210; meanwhile the mean ACT score is 21 with a standard deviation of 5.4. A student takes both tests and receives a 1350 on the SAT and a 29 on the ACT. On which test did the student score better?

Percentiles

Percentile

A **percentile** is a measure of location that divide a set of *ordered* data into 100 groups with about 1% of the values in each group.

Percentile Score

A **percentile score** is the percent of data values less than a given value. (*Note*: this is not the same as percentage).

Example 2. Explain the difference between getting 90% on a test and scoring in the 90th percentile on that test.

Note: When working with percentiles, it is important to know whether or not the data value in question is included in the calculation.

Example 3. Determine the percentile score value for the test score of 85 in the following data set. Only include values that are less than 85 in your calculation.

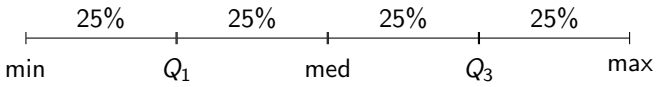
74, 74, 76, 77, 83, 85, 85, 90, 93, 94, 97, 98

Example 4. Determine the percentile score for 85 when you include values less than *or equal to* 85.

Five-Number Summary

Quartiles

Quartiles are values that divide a data set into 4 groups, with each group holding 25% of the data.



Q_1 is called the first (a.k.a. *lower*) quartile and Q_3 is called the third (a.k.a. *upper*) quartile.

Five-Number Summary

The **five-number summary** are the following values:

Minimum, Q_1, Median, Q_3, Maximum

Example 5. Find the five-number summary of the following data set:

1, 2, 2, 4, 5, 7, 11, 15, 15, 18, 44

Detecting Outliers

Outliers

An outlier is an extreme data value in a dataset.

We can use the five-number summary to detect outliers.

Interquartile Range

The **interquartile range** can be found by subtracting Q_1 from Q_3:

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

The **lower fence** is

$$Q_1 - 1.5(\text{IQR})$$

and the **upper fence** is

$$Q_3 + 1.5(\text{IQR})$$

A data value is an **outlier** if it is less than the lower fence or more than the upper fence.

Example 6. Calculate the lower and upper fences of the previous example's data set and use it to find any outliers.

Box Plots

Boxplot

A **box plot** (a.k.a. box-and-whisker plot) is a visual display of the five-number summary.

Example 7. Create a box plot for the data set

1, 2, 2, 4, 5, 7, 11, 15, 15, 18, 44

Modified Box Plot

- Outliers are shown with symbols such as stars or points.
- Whiskers are drawn out to the points in the data set that are **not** considered outliers.