

AP Statistics Unit 1 Overview

Exploring One-Variable Data

AP Statistics Unit 1: Exploring One-Variable Data is all about understanding and interpreting data relating to one variable!

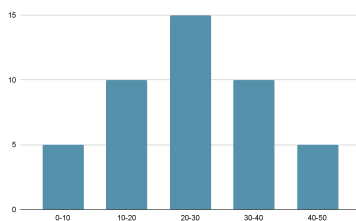
Describing Distribution: When given data, it is important to be able to understand the features relating to the data. Describing distributions can help us understand the data in context.

There are 4 ways to describe data distributions:

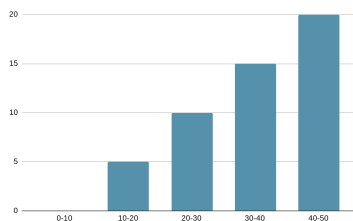
- Shape
- Outliers
- Center
- Spread

The acronym is S.O.C.S.!

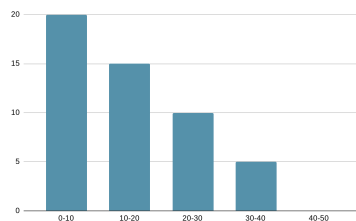
The **shape** describes the appearance of the data on a stem and leaf plot, dot plot, or histogram, and it shows where most of the data lies.



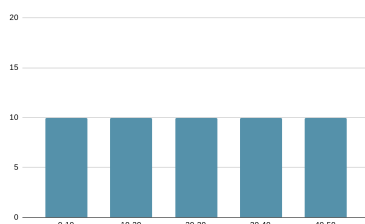
NORMAL



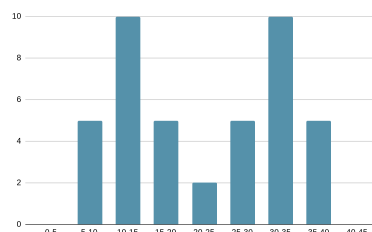
LEFT SKEWED



RIGHT SKEWED



UNIFORM



BIMODAL

The **outliers** tell us if there's any unusual data! An outlier is very different from the rest of the data, and we can use the IQR test to determine if something is an outlier.

We know that a piece of data is an outlier if it's more than 1.5 IQR away from Q1 or Q3 (see below!)

The **center** can be represented in 2 ways!

- The **mean** tells us the average of the data
- The **median** tells us the middle of the data

In data with outliers, using the median could give us a better sense of the true middle. If the data doesn't have outliers, feel free to use the mean!

The **spread** or variation tells us about the layout of our data. Is the data all very close together? Is it spread out?

- The **range** is the difference between the maximum and minimum
- The **IQR** is the difference between the Q3 and Q1 data, the 2 innermost quartiles
 - If we had 100 pieces of data, the IQR would be data #75 - data #25

Types of Data: There are different types of data!

Categorical	Quantitative	Discrete	Continuous
This is a type of data that is collected in categories, <i>not</i> numbers! Like color or food preference!	This is a type of data that is collected in numbers, and can be measured in some way! Like age or points scored!	This is a type of data that doesn't have infinite possibilities between numbers. Like grades in school!	This is a type of data that does have infinite possibilities between numbers. Like time taken to complete an event!
Food, school subject, color, movie, etc.	Age, points scored, weight, height, etc.	People, objects, animals, successes/failure, etc.	Time to complete an event, age, speed/velocity, length, etc.

The **standard deviation** of a set of data tells us the average difference between a piece of data and the mean of the data set. It can indicate to us how the data is spread. (i.e. Data with a large standard deviation is more spread than data with a small standard deviation.)

The **z-score** of a certain piece of data tells us how unusual this piece of data is, while assuming a normal + symmetric distribution.

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

z = z-score

x = individual data value

μ = dataset mean

σ = dataset standard deviation

