

9.2 Data Distribution

A Practice Understanding Task



A lot of information can be obtained from looking at data plots and their distributions. It is important when describing data that we use context to communicate the **shape, center, and spread**.

Shape and spread:

- **Modes:** uniform (evenly spread- no obvious mode), unimodal (one main peak), bimodal (two main peaks), or multimodal (multiple locations where the data is relatively higher than others).
- **Skewed distribution:** when most data is to one side leaving the other with a 'tail'. Data is skewed to side of tail. (if tail is on left side of data, then it is skewed left).
- **Normal distribution and standard deviation:** curve is unimodal and symmetric. Data that has a normal distribution can also describe the data by how far it is from the mean using standard deviation.
- **Outliers:** values that stand away from the body of the distribution. For a box-and-whisker outliers determined if they are more than 1.5 times the interquartile range (length of box) beyond quartiles 1 and 3. Also considered an outlier if data is more than two standard deviations from the center of a normal distribution.
- **Variability:** values that are close together have low variability; values that are spread apart have high variability.

Center:

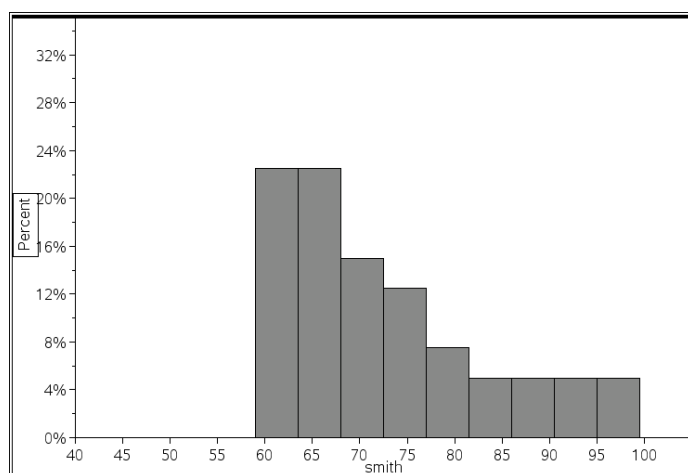
- Analyze the data and see if one value can be used to describe the data set. Normal distributions make this easy. If not a normal distribution, determine if there is a 'center' value that best describes the data. Bimodal or multimodal data may not have a center that would provide useful data.

There are representations of test scores from six different classes found below, for each:

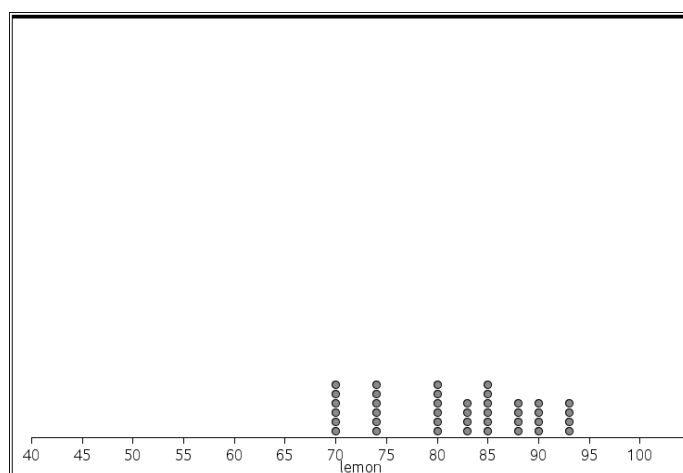
1. Describe the data distribution.
2. Compare data distributions between Anderson and Williams.
3. Compare data distributions between Williams and Lemon.
4. Compare data distributions between Croft and Hurlea.
5. Compare data distributions between Jones, Spencer, and Anderson.
6. Compare data distributions between Spencer and the other histograms.
7. Which distributions are most similar? Different? Explain your answer.

SECONDARY MATH I // MODULE 3
MODELING DATA—9.2

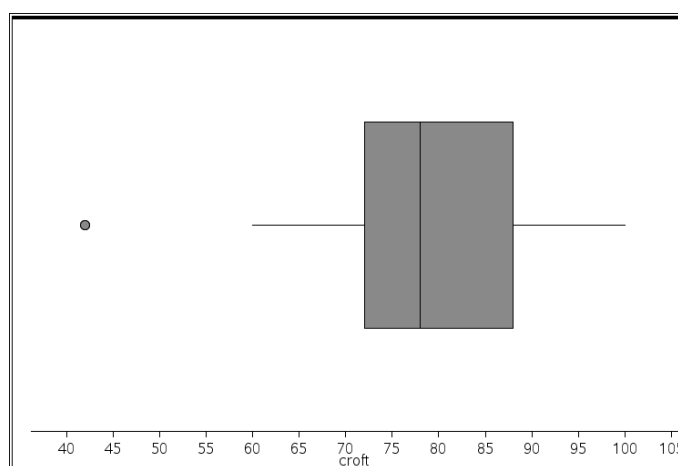
Data set I: Williams's class



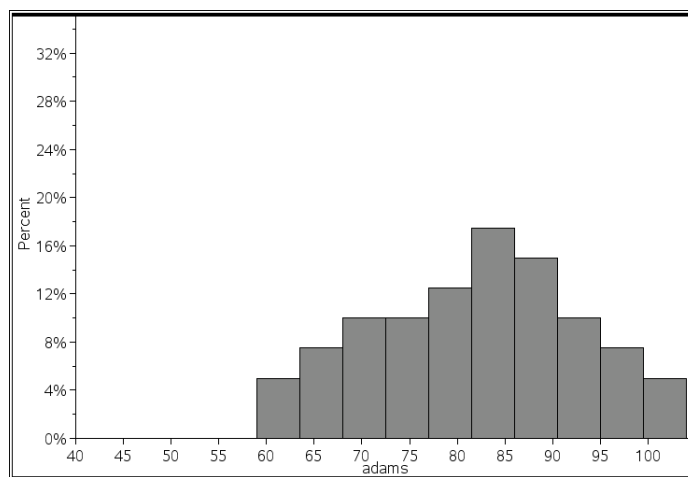
Data set II: Lemon's class



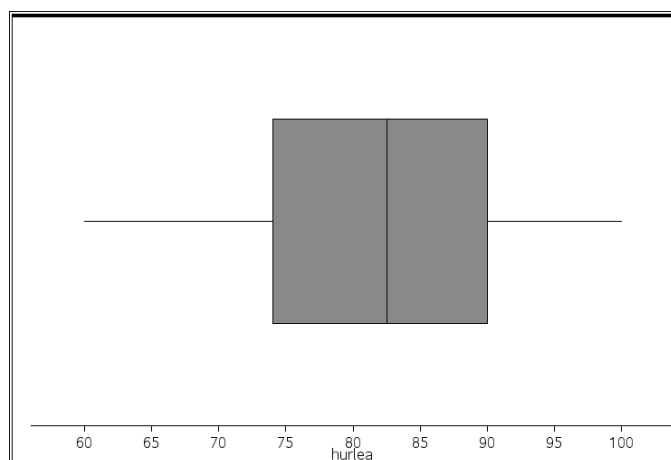
Data set III: Croft's Class



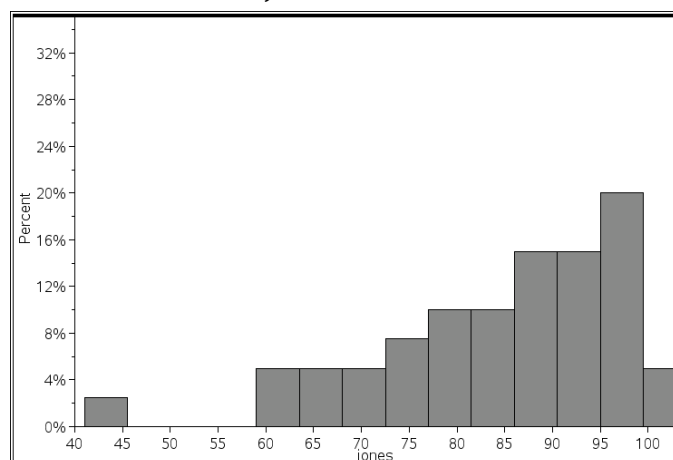
Data set IV: Anderson's Class



Data set V: Hurlea's class

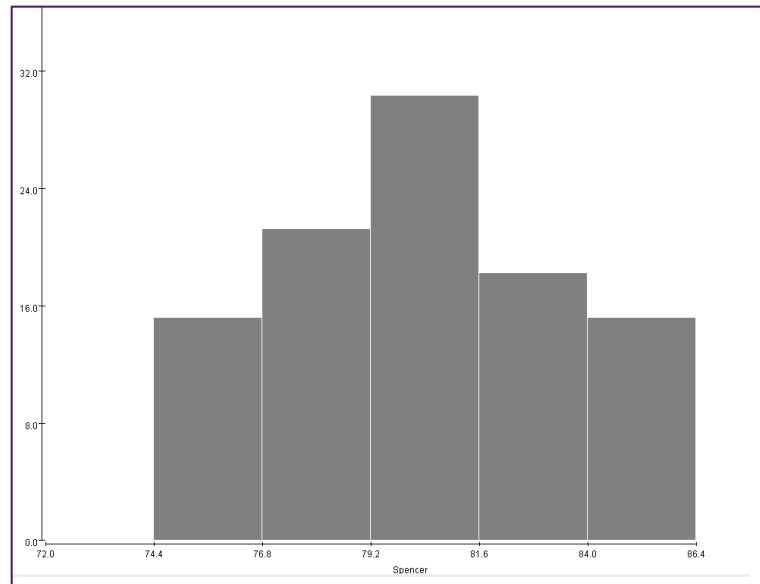


Data set VI: Jones' class



SECONDARY MATH I // MODULE 3
MODELING DATA–9.2

Data set VII: Spencer's class



Data set VIII: Overall Achievement Test Scores

