

Density Curves, Normal Curve, & 68-95-99.7 Rule

Exploring Quantitative Data

1.

2.

3.

4.

Density Curve

A **density curve**:

It has a few key features:

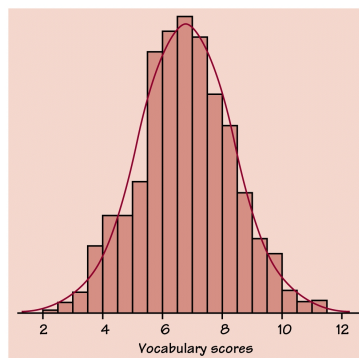
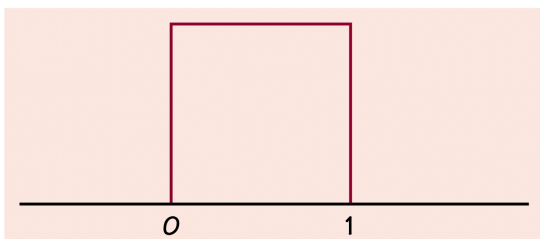
1.

2.

The densities (areas) under the curve bounded by values represent proportions (or relative frequencies) of the data in the distribution between those values.

Density curve with uniform distribution

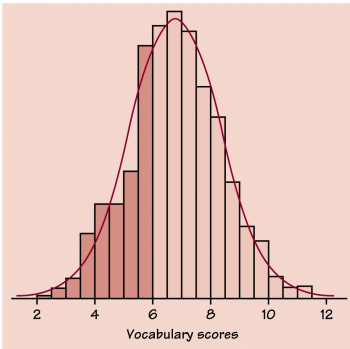
Density curve fit to histogram



Example: Vocabulary Scores

To find the proportion of vocabulary scores between 2 and 6:

Compare the method using the actual histogram below to the method with a density curve that fits the distribution of the histogram:



Histogram method: Add all the frequencies in the bins between 2 and 6. Divide by the total number of frequencies.

Density curve method: Calculate the density (area) under the curve between 2 and 6.

(Note: We haven't done this yet, but it's fast and accurate, even when the data is unknown. That's because every density curve is modeled by a density function.)

Density Curves: Shape and Center

Mean of a density curve:

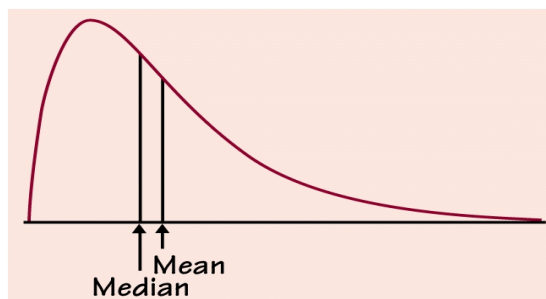
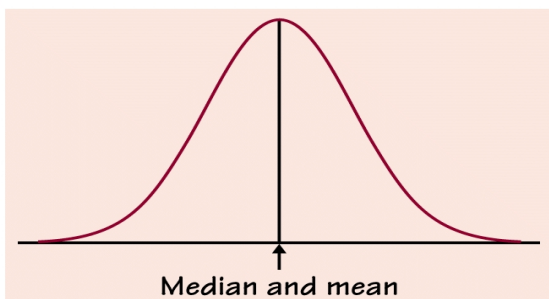
Median of a density curve:

Mode of a density curve:

Unimodal:

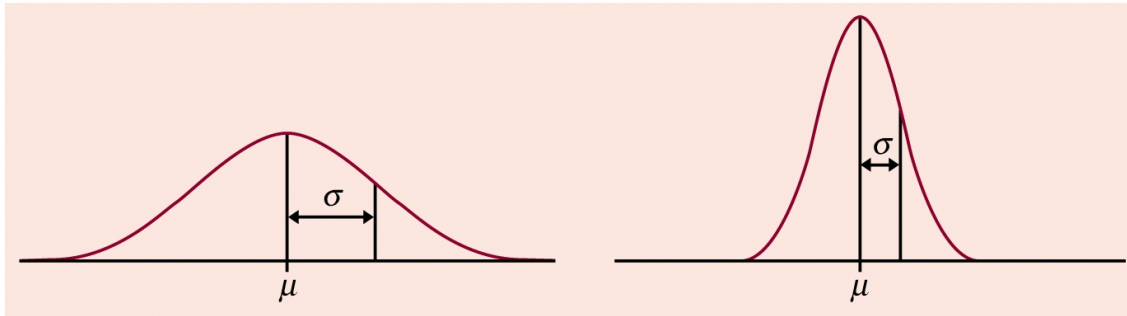
A very common shape of density curve is a unimodal distribution. In a density curve, the mode is always the highest point (peak) of the curve. The median and mean are the same for a symmetric density curve. They both lie at the center of the curve. The mean of a skewed curve is pulled away from the median in the direction of the long tail.

A unimodal density curve can be symmetric in shape or skewed to the left or right.



Normal Density Curves:

The Normal density curve (also called a Gaussian distribution) is the most famous and important of all density curves.



The Normal density curve is actually an entire class (or infinite set) of density curves.

μ

σ

(Note: Greek letters are used for the mean and standard deviation because the density curve is a mathematical model. \bar{x} and s_x are used to indicate the mean and standard deviation of the actual data in a distribution)

We use the notation $N(\mu, \sigma)$ to define a specific Normal density curve.

All Normal density curves have the following properties:

1.

2.

3.

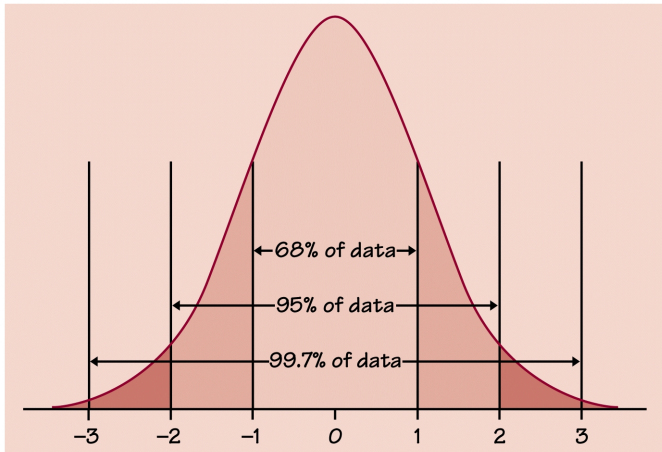
4.

5.

68-95-99.7 Rule

For all Normal density curves:

- Approximately 68% of the density (area) is within σ of μ
- Approximately 95% of the density (area) is within 2σ of μ
- Approximately 99.7% of the density (area) is within 3σ of μ



Example: Using the 68-95-99.7 Rule

The scores for a particular test have a distribution $N(80, 5)$.

1. Sketch a picture of the density curve for this distribution.
2. Label the value at μ and the values at $\mu \pm \sigma$, $\mu \pm 2\sigma$, and $\mu \pm 3\sigma$.
3. Answer the questions below.

What proportion of the test scores are more than 90?

What proportion of the test scores are between 70 and 85?

What proportion of the test scores are less than 65?

If an outlier is more than three standard deviations from the mean, what proportion of test scores are outliers?

