

# MAST7866

## Shapes of Distributions

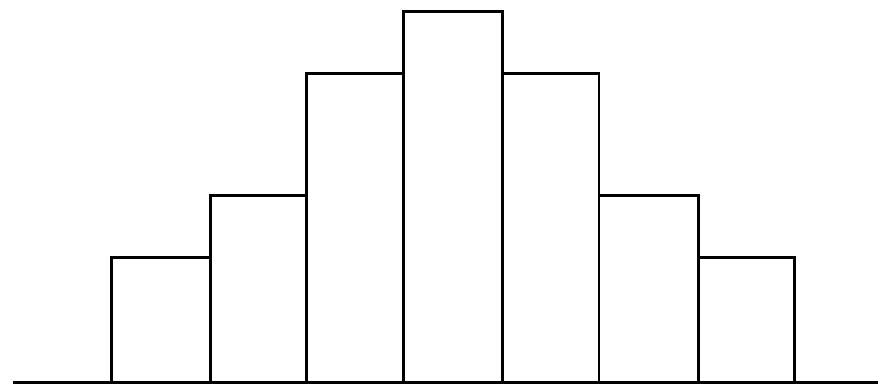
Jian Zhang  
University of Kent

# Shapes of Distributions

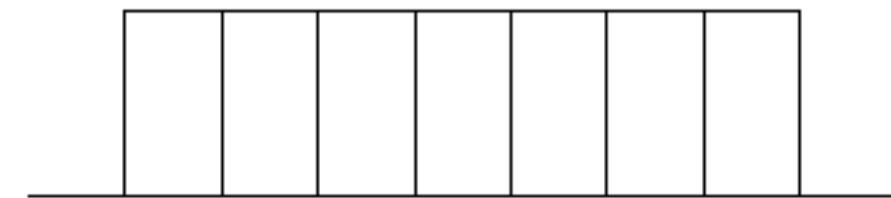
- If a sample is representative of a population, then a histogram of the sample data should have a shape that is similar to **the distribution of a population**.

# Common Shapes of Distributions

Symmetric

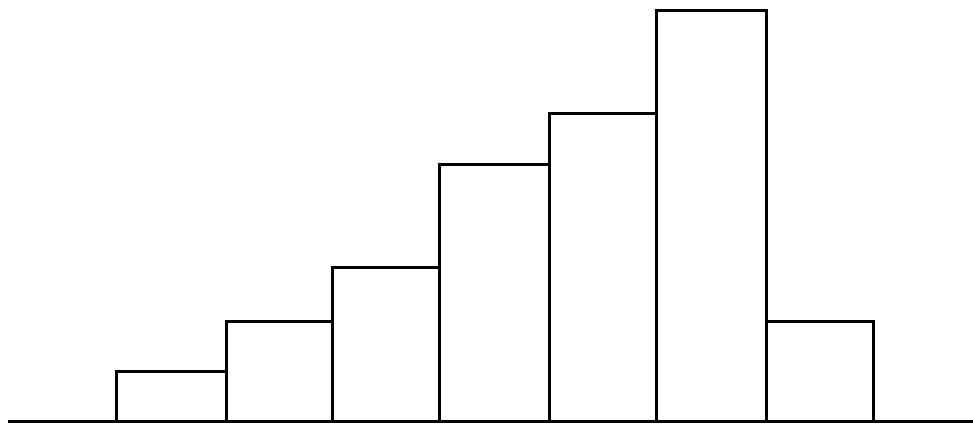


Uniform

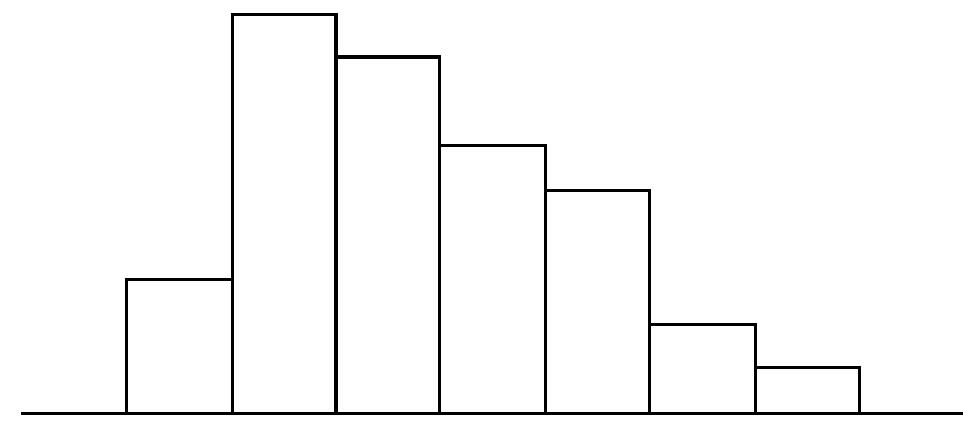


# Shapes of Distributions (skewness)

Negatively skewed

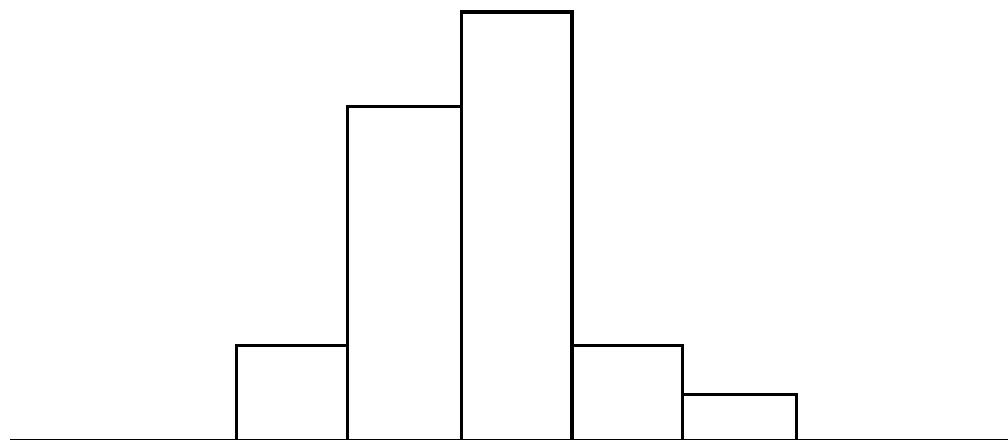


Positively skewed

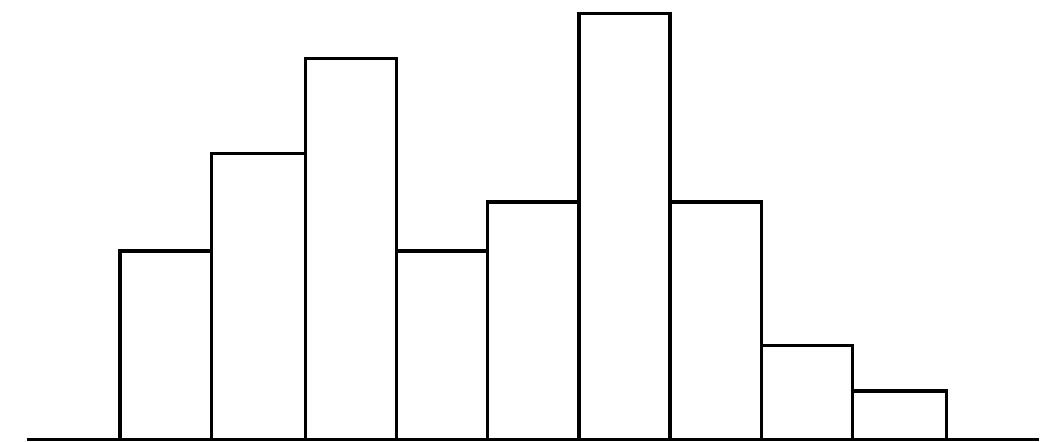


# Shapes of Distributions (unimodal/bimodal)

Unimodal

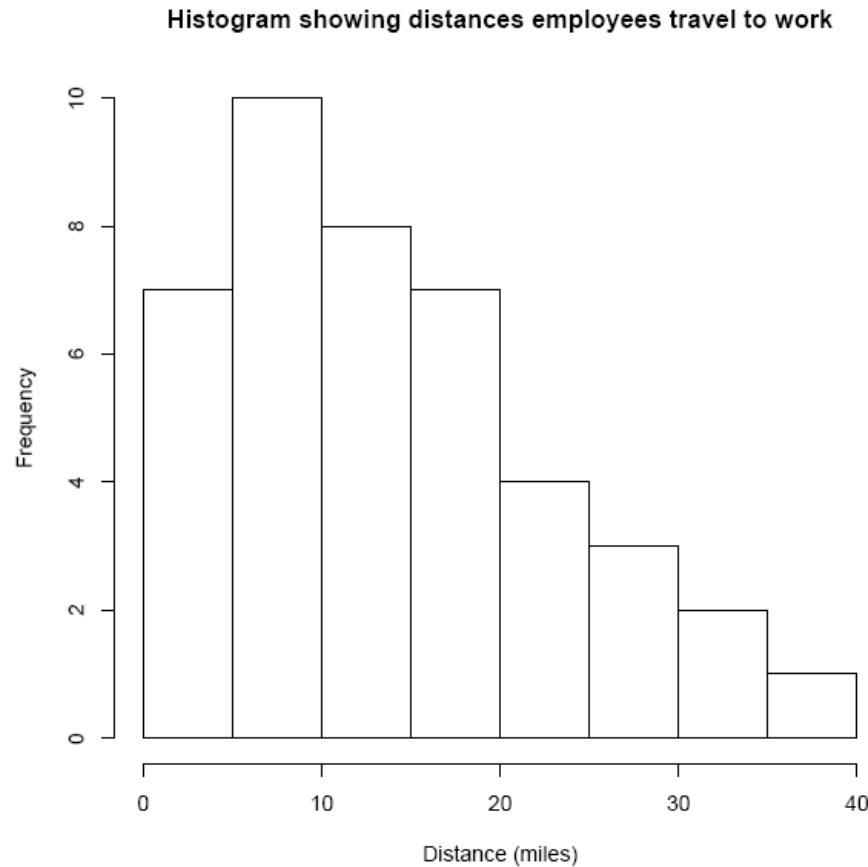


Bimodal



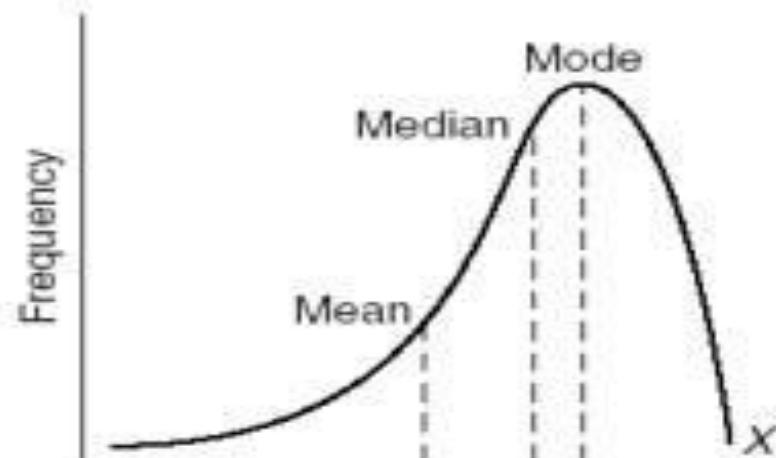
# Travel to work distances

What shape does the distribution of distances employees travel to work considered shown here have?

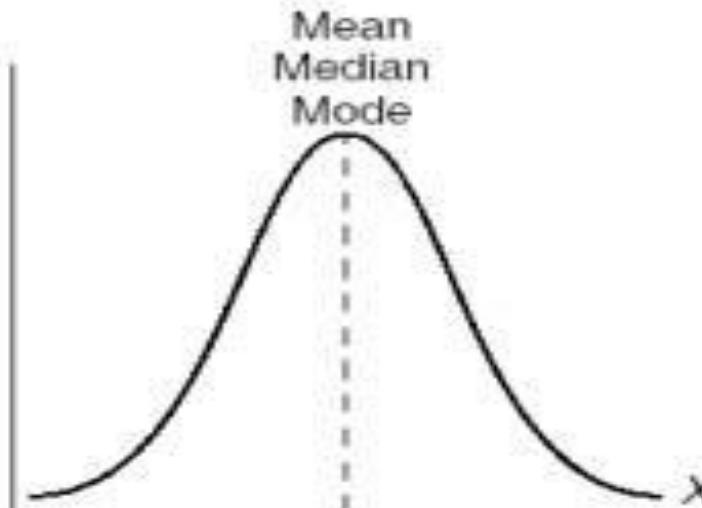


# Relationships between shapes/measures

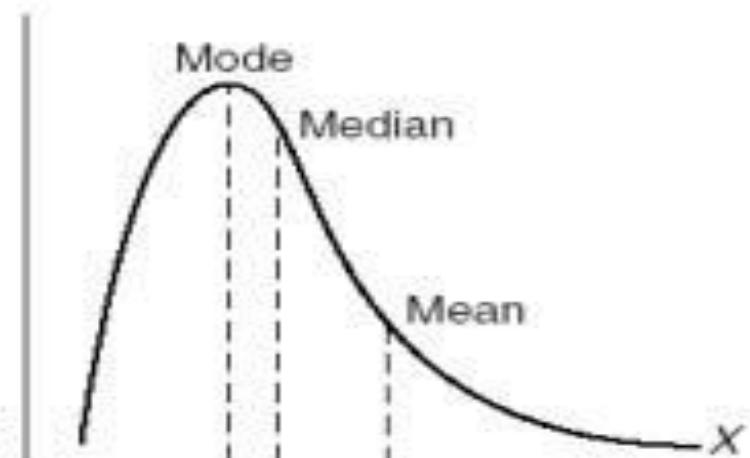
(a) Negatively skewed



(b) Normal (no skew)



(c) Positively skewed

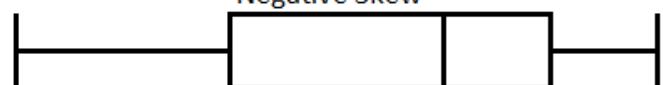


Negative direction

The normal curve  
represents a perfectly  
symmetrical distribution

Positive direction

Negative Skew



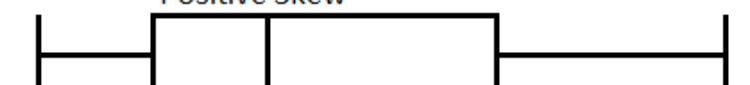
Negatively skewed distribution  
 $\text{mean} < \text{median} < \text{mode}$

symmetrical



Symmetric distribution  
 $\text{mean} = \text{median} = \text{mode}$

Positive Skew



Positively skewed distribution  
 $\text{mean} > \text{median} > \text{mode}$

# Skewness

**Skewness** is a measure of symmetry, or more precisely, the **lack of symmetry**.

A distribution, or data set, is symmetric if it looks the same to the left and right of the centre point.

# Interpreting Skewness

How to interpret skewness?

- The skewness for a symmetric distribution is zero.
- Negative values indicate data that are skewed left – i.e., have left tails which are long relative to the right tail (negatively skewed)
- Positive values indicate data that are skewed right – i.e., have right tails which are long relative to the left tail (positively skewed)

# Kurtosis

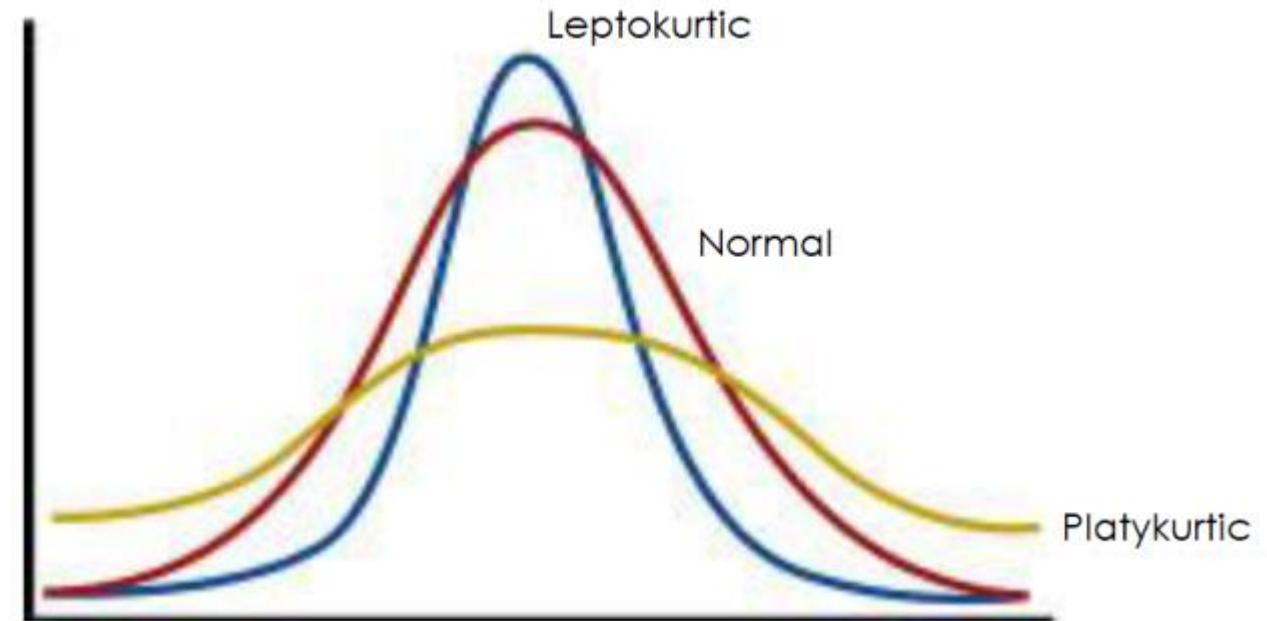
**Kurtosis** is a measure of whether the data are **heavy-tailed** or **light-tailed** relative to a normal distribution.

That is, data sets with high kurtosis tend to have heavy tails, or outliers.

Data sets with low kurtosis tend to have light tails, or lack of outliers. A uniform distribution would be the extreme case.

# Visualising Kurtosis

- The kurtosis for a standard normal distribution is 3.
- A positive excess kurtosis indicates a “heavy-tailed” distribution – also referred to as **Platykurtic**
- A negative excess kurtosis indicates a “light-tailed” distribution – also referred to as **Leptokurtic**



<https://javashine.wordpress.com/2017/08/21/measures-of-shape-skew-kurtosis/>

# Obtaining moments in R

- In R you will need to install the package moments:

```
> install.packages("moments")  
> library(moments)  
> skewness(data)  
> kurtosis(data)
```