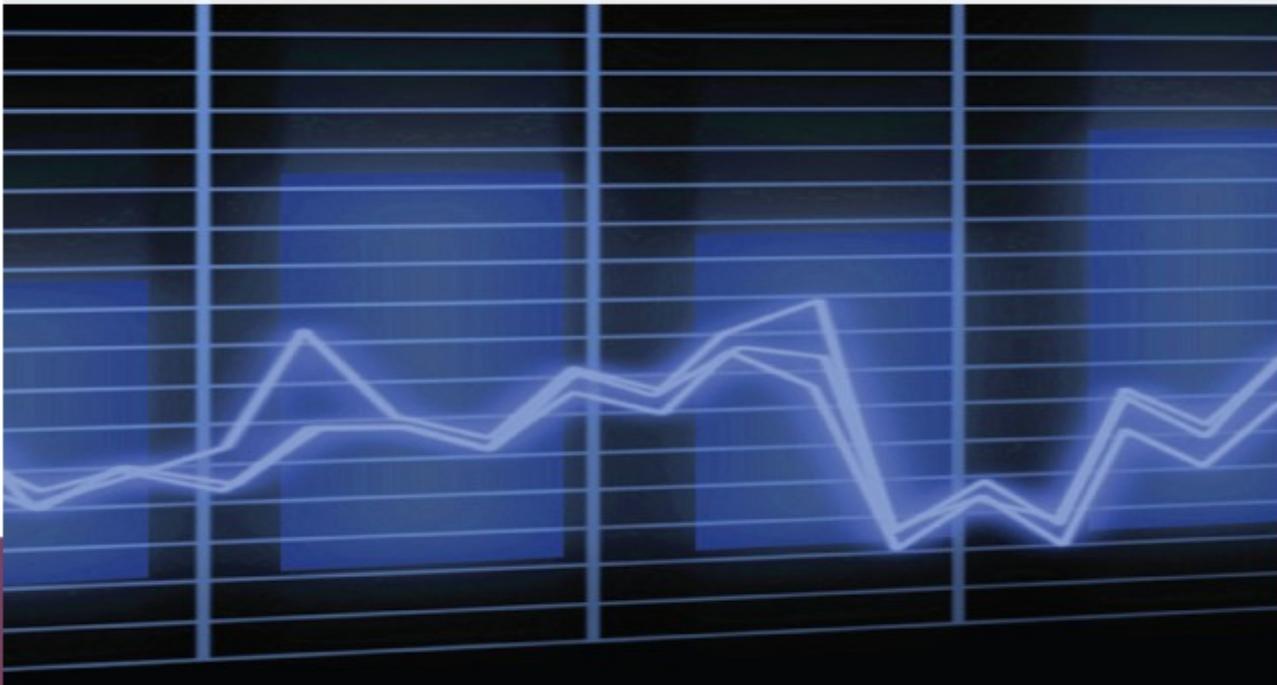


# Statistics



# 4.1

## Population, Sample, and Data



# Histograms and Relative Frequency Density

# Histograms and Relative Frequency Density

Density is a ratio. In science, density is used to determine the concentration of weight in a given volume:

Density = weight/volume.

For example, the density of water is 62.4 pounds per cubic foot. In statistics, density is used to determine the concentration of data in a given interval:

Density = (percent of total data)/(size of an interval).

Because relative frequency is a measure of the percentage of data within an interval, we shall calculate the relative frequency density of an interval to determine the concentration of data within the interval.

# Histograms and Relative Frequency Density

For example, if the interval  $20 \leq x < 25$  contains eleven out of forty-two data points, then the relative frequency density of the interval is

$$rfd = \frac{f/n}{\Delta x} = \frac{11/42}{5} = 0.052380952 \dots$$

## Definition of Relative Frequency Density

Given a set of  $n$  data points, if an interval contains  $f$  data points, then the **relative frequency density** ( $rfd$ ) of the interval is

$$rfd = \frac{f/n}{\Delta x}$$

where  $\Delta x$  is the width of the interval.

# Histograms and Relative Frequency Density

If a histogram is constructed using relative frequency density as the vertical scale, the area of a rectangle will correspond to the relative frequency of the interval, as shown in Figure 4.7.

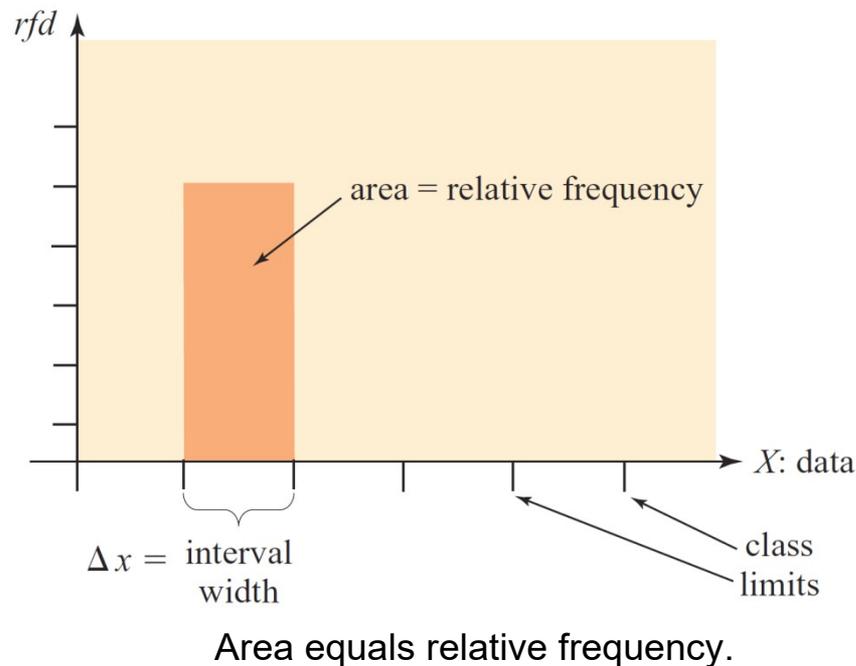


Figure 4.7

# Histograms and Relative Frequency Density

area = base  $\cdot$  height

$$= \Delta x \cdot rfd$$

$$= \Delta x \cdot \frac{f/n}{\Delta x}$$

$$= f/n$$

= relative frequency

# Histograms and Relative Frequency Density

Adding a new column to the frequency distribution given in Figure 4.5, we obtain the relative frequency densities shown in Figure 4.8.

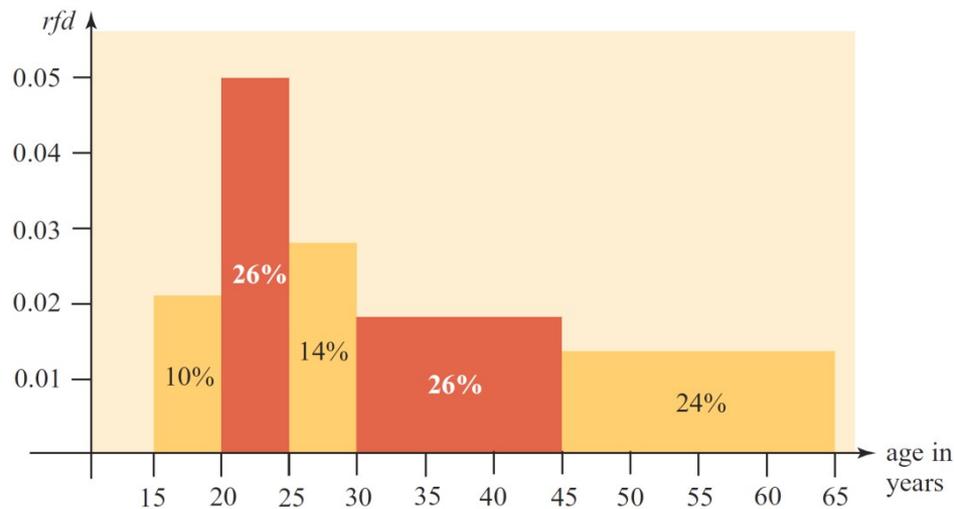
$x = \text{Age (years)}$	Relative Frequency	Class Frequency	Relative Width	Frequency Density
$15 \leq x < 20$	4	$\frac{4}{42} \approx 10\%$	5	$\frac{4}{42} \div 5 \approx 0.020$
$20 \leq x < 25$	11	$\frac{11}{42} \approx 26\%$	5	$\frac{11}{42} \div 5 \approx 0.052$
$25 \leq x < 30$	6	$\frac{6}{42} \approx 14\%$	5	$\frac{6}{42} \div 5 \approx 0.028$
$30 \leq x < 45$	11	$\frac{11}{42} \approx 26\%$	15	$\frac{11}{42} \div 15 \approx 0.017$
$45 \leq x < 65$	10	$\frac{10}{42} \approx 24\%$	20	$\frac{10}{42} \div 20 \approx 0.012$
	$n = 42$	Total = 100%		

Calculating relative frequency density.

Figure 4.8

# Histograms and Relative Frequency Density

We now construct a histogram using relative frequency density as the vertical scale. The histogram depicting the distribution of the ages of the people in Keith Reed's acting class (using the frequency distribution in Figure 4.8) is shown in Figure 4.9.

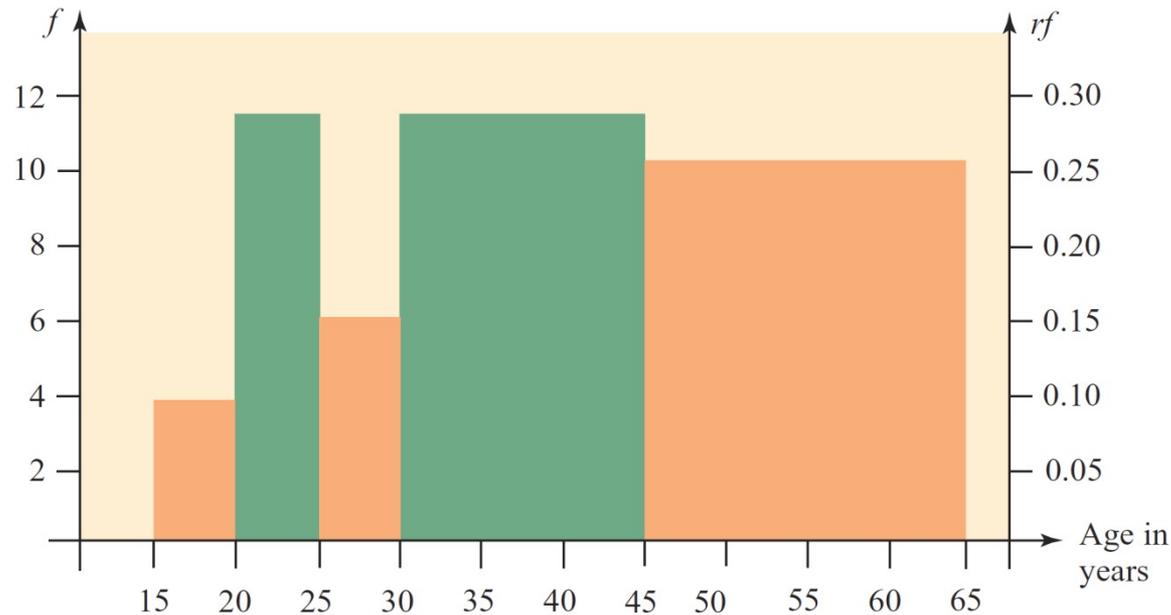


Ages of the people in Keith Reed's acting class.

Figure 4.9

# Histograms and Relative Frequency Density

Comparing the histograms in Figures 4.6 and 4.9, we see that using relative frequency density as the vertical scale (rather than frequency) gives a more truthful representation of a distribution when the interval widths are unequal.



Why is this histogram misleading?

Figure 4.6

## Example 3 – *Constructing a Histogram: Grouped Data*

To study the output of a machine that fills bags with corn chips, a quality control engineer randomly selected and weighed a sample of 200 bags of chips. The frequency distribution in Figure 4.10 summarizes the data. Construct a histogram for the weights of the bags of corn chips.

$x = \text{Weight (ounces)}$	$f = \text{Number of Bags}$
$15.3 \leq x < 15.5$	10
$15.5 \leq x < 15.7$	24
$15.7 \leq x < 15.9$	36
$15.9 \leq x < 16.1$	58
$16.1 \leq x < 16.3$	40
$16.3 \leq x < 16.5$	20
$16.5 \leq x < 16.7$	12

Weights of bags of corn chips.

Figure 4.10

## Example 3 – Solution

Because each interval has the same width ( $\Delta x = 0.2$ ), we construct a combined frequency and relative frequency histogram. The relative frequencies are given in Figure 4.11.

$x$	$f$	$rf = f/n$
$15.3 \leq x < 15.5$	10	0.05
$15.5 \leq x < 15.7$	24	0.12
$15.7 \leq x < 15.9$	36	0.18
$15.9 \leq x < 16.1$	58	0.29
$16.1 \leq x < 16.3$	40	0.20
$16.3 \leq x < 16.5$	20	0.10
$16.5 \leq x < 16.7$	12	0.06
	$n = 200$	Sum = 1.00

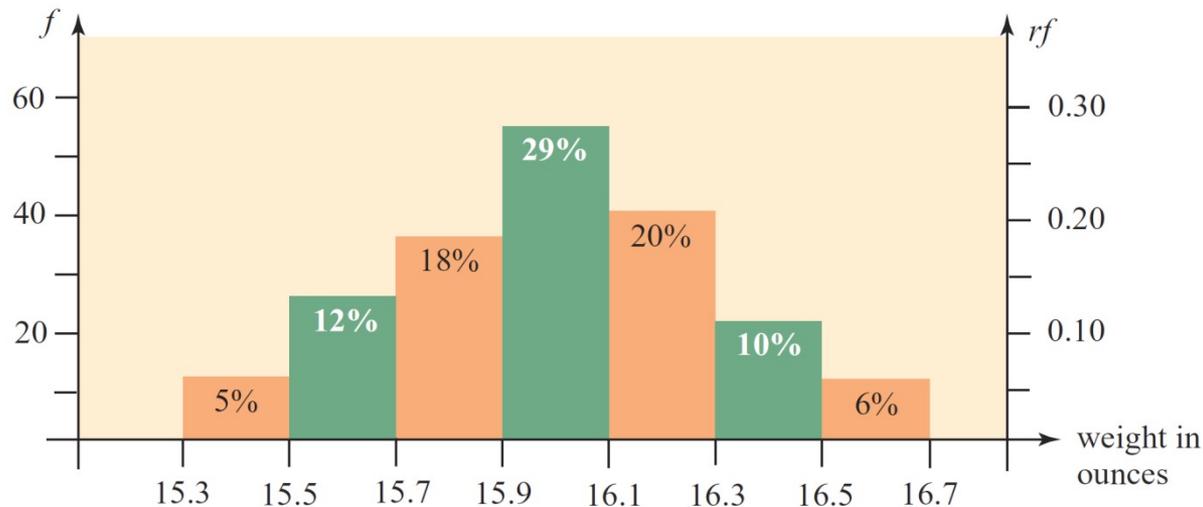
Relative frequencies.

Figure 4.11

# Example 3 – Solution

cont'd

We now draw coordinate axes with appropriate scales and rectangles (Figure 4.12). Notice the (near) symmetry of the histogram.



Weights of bags of corn chips.

Figure 4.12