Probability

Continuous Random Variable

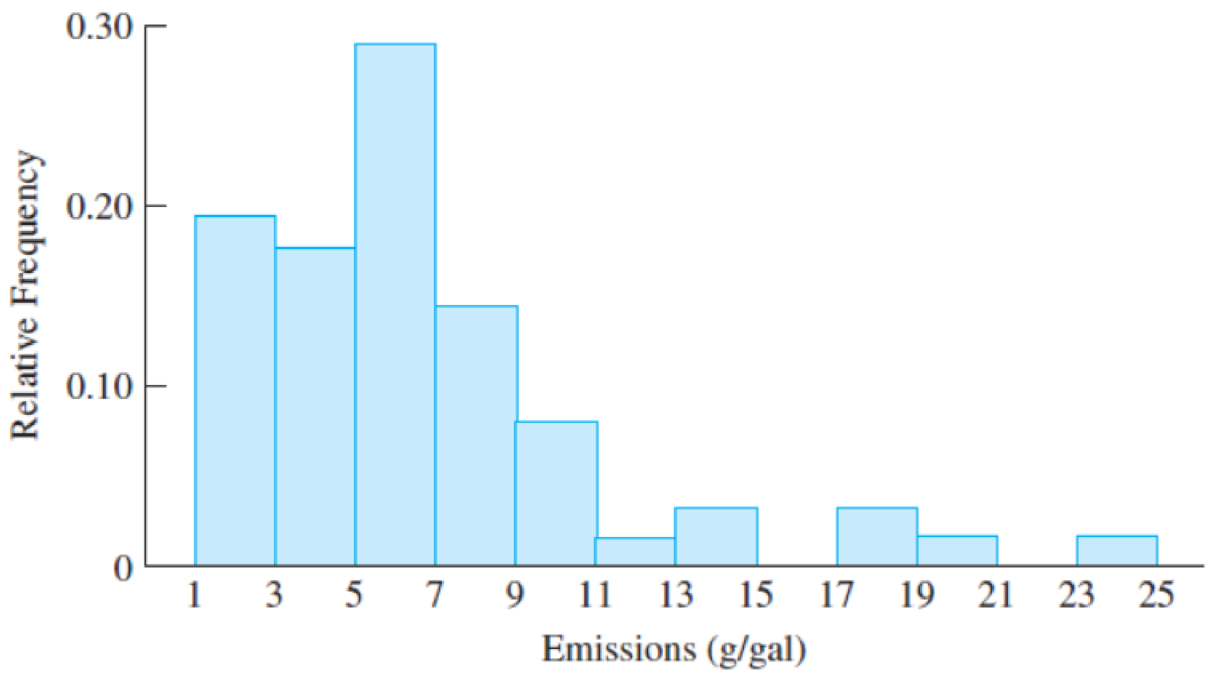
Prof. Dr. Mostafa Elhosseini

https://youtube.com/drmelhosseini

Histogram

- A **histogram** is a graphic that gives an idea of the **shape** of a sample, indicating **regions** where sample points are **concentrated** and regions where they are sparse
- **Frequency** presents the **numbers of data points** that fall into each of the class intervals
- **Relative Frequency** presents the frequencies divided by the total number of data points
- since every data point is in exactly one class interval, the **relative frequencies must sum to 1**
- **Density** presents the **relative frequency divided by the class width**

Histogram



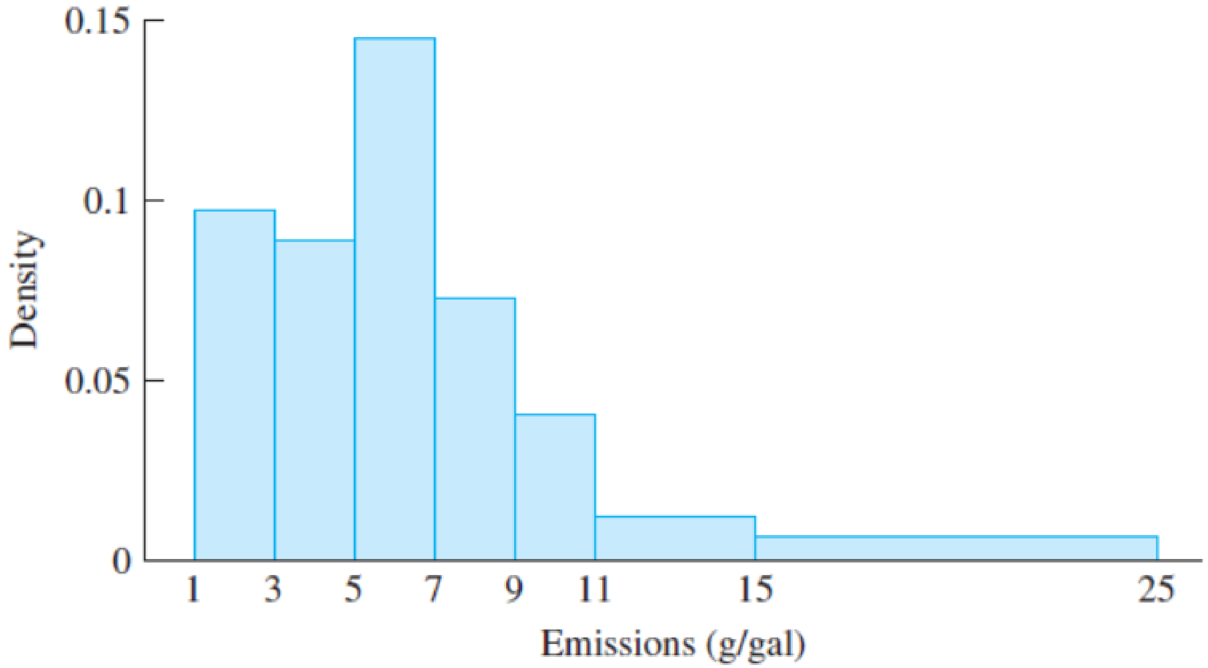
- **In some cases**, histograms are drawn with **class intervals** of **differing** widths.

Histogram – unequal class width

Class Interval (g/gal)	Frequency	Relative Frequency	Density
1–< 3	12	0.1935	0.0968
3–< 5	11	0.1774	0.0887
5–< 7	18	0.2903	0.1452
7–< 9	9	0.1452	0.0726
9–< 11	5	0.0806	0.0403
11–< 15	3	0.0484	0.0121
15–< 25	4	0.0645	0.0065

- Because the **class widths vary in size**, the **densities** are **no** longer **proportional to the relative frequencies**. Instead, the densities adjust the relative frequency for the width of the class.

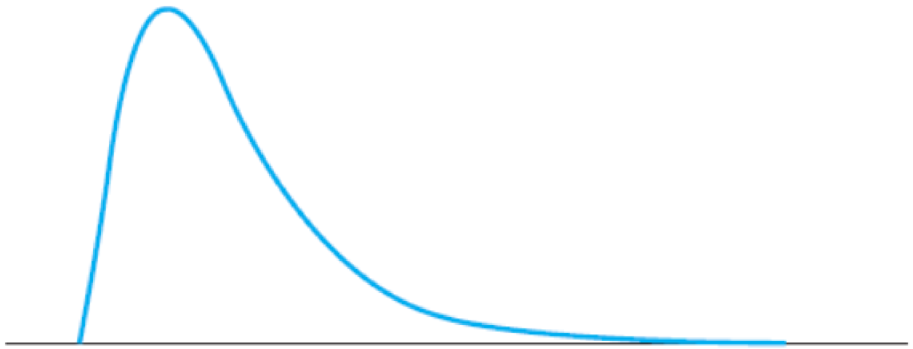
Histogram – unequal class width



- when the **classes have unequal widths**, the **heights** of the rectangles must be **set equal to the densities**.
- The **areas** of the rectangles then **represent the relative frequencies**

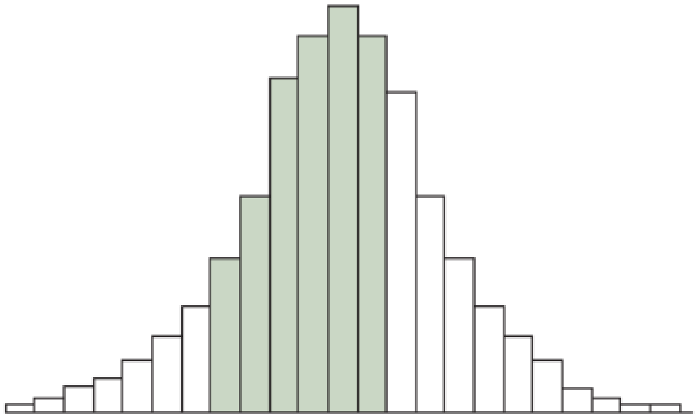
Continuous Random Variable

- Note that **emissions is a continuous variable**, because its **possible values** are **not restricted to some discretely spaced set**
- if we had information on the entire population, **containing millions of vehicles**, we could make the **intervals extremely narrow**. The histogram would then look quite smooth and could be approximated with a curve



Probability Density Function (PDF)

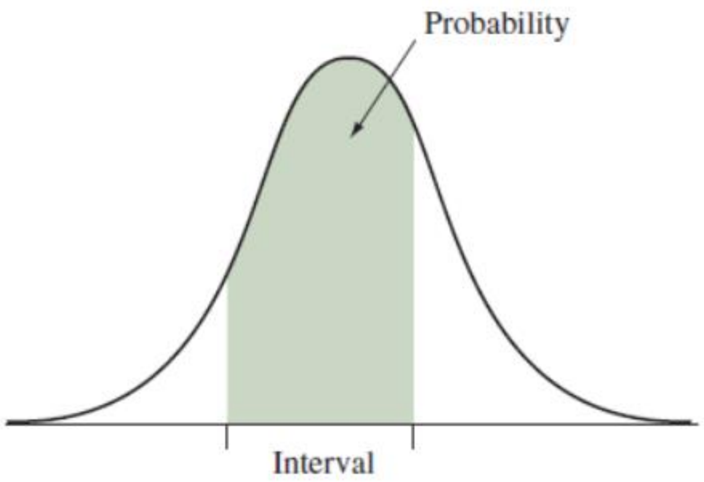
- As the number of intervals increases, with their width narrowing, the shape of the histogram gradually approaches a smooth curve. We'll use such curves to portray probability distributions of continuous random variables



Many intervals

Probability Density Function (PDF)

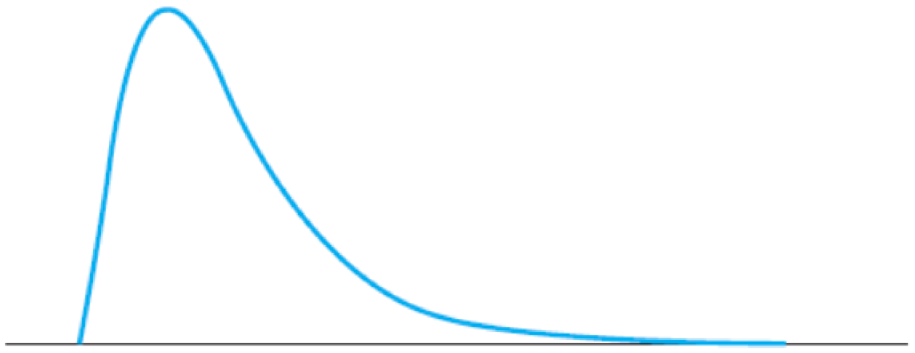
- A **continuous random variable** has possible values that form an interval. Its probability distribution is specified by a curve that determines the probability that the random variable falls in any particular interval of values.
- **Each interval has probability between 0 and 1**. The probability for the interval is given by the area under the curve, above that interval.
- The interval containing all possible values has probability equal to 1, so the total **area under the curve equals 1**.



Smooth curve approximation

Continuous Random Variable

- If a vehicle were chosen at random from this population to have its emissions measured, the emissions level X would be a random variable.
- The **probability that X falls between any two values a and b** is equal to the **area** under the histogram between a and b
- Because the histogram in this case is represented by a **curve**, the **probability** would be found by computing an **integral**



Probability Density Function (PDF)

- A **continuous random variable** is defined to be a random variable whose probabilities are represented by **areas under a curve**.
- This curve is called the **probability density function**
- Let $\boldsymbol{X}$ be a continuous random variable. Let the function $\boldsymbol{f(x)}$ be the probability density function of $\boldsymbol{X}$. Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be any two numbers, with $\boldsymbol{a} < \boldsymbol{b}$.
- The proportion of the population whose values of X lie between a and b is given by $\int_{\boldsymbol{a}}^{\boldsymbol{b}} \boldsymbol{f(x)\,dx}$ the area under the probability density function between a and b

Probability Density Function (PDF)

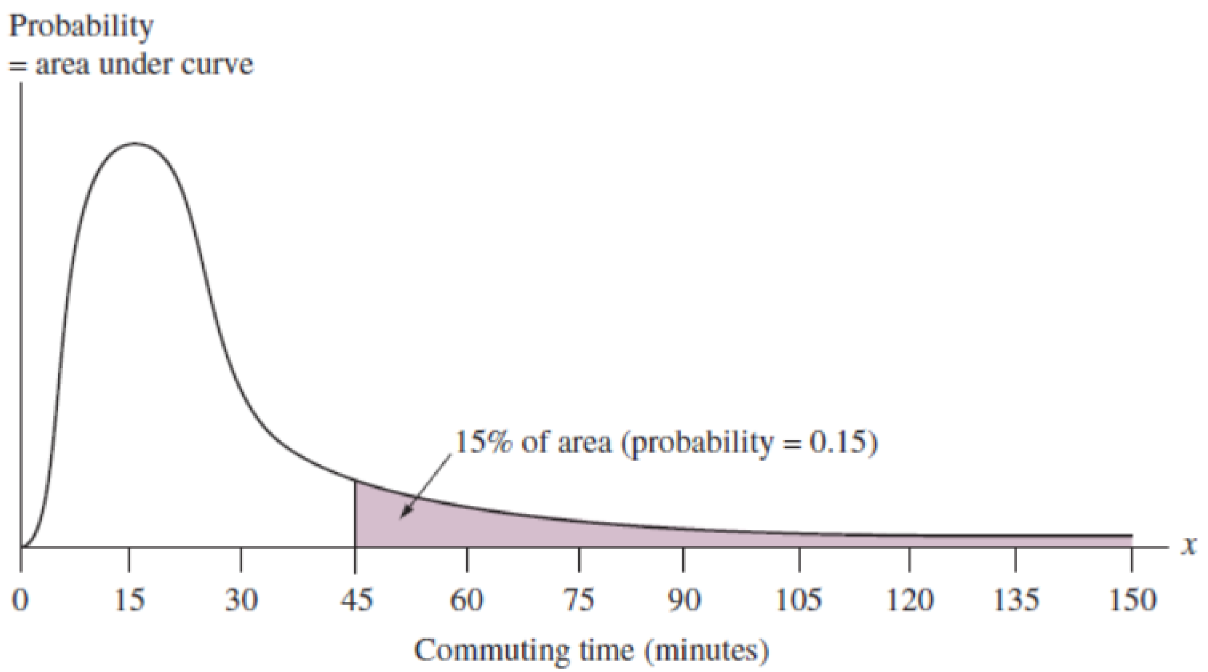
- The proportion of the population whose values of X lie between a and b is given by

$$\int_{a}^{b} \boldsymbol{f(x)\,dx}$$

- Probability that the random variable X takes on a value between a and b.
- Note that the **area under the curve does not depend on** whether the endpoints a and b are included in the interval

Example

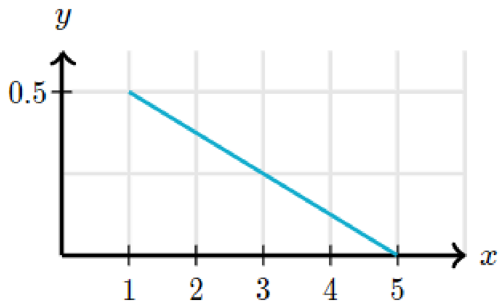
- Probability distribution of X = **commuting time** to work for workers in the United States, based on information from the 2009 American Community Survey



▲ **Figure 6.2 Probability Distribution of Commuting Time.** The area under the curve for values higher than 45 is 0.15. **Question** Identify the area under the curve representing the probability of a commute less than 15 minutes, which equals 0.27.

- New book on Stat
- Khan academy – my notes
- Rule – normal distribution

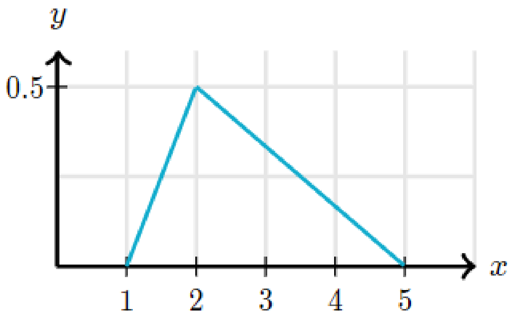
Consider the density curve below.



Find the probability that x is more than 3.

$P(x > 3) =$ ______

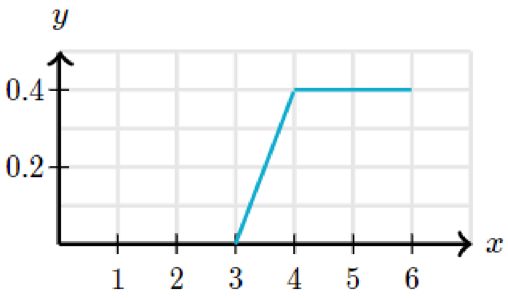
Consider the density curve below.



Find the probability that x is greater than 2.

$P(x > 2) =$ ______

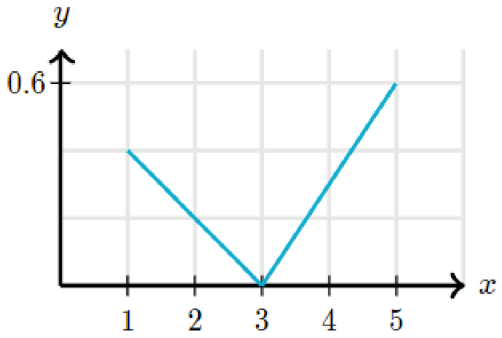
Consider the density curve below.



Find the probability that x is more than 3.

$P(x > 3) =$ ______

Consider the density curve below.



Find the probability that x is less than 3.

$P(x < 3) =$ ______

References

- Statistics for Engineers and Scientists, Fourth Edition, William Navidi Colorado School of Mines
- Agresti, A., Franklin, C. and Klingenberg, B., 2017. Statistics: The art and science of learning from data (Fourth Edi).
- Khan Academy