Probability

Continuous Random Variable

Prof. Dr. Mostafa Elhosseini

https://youtube.com/drmelhosseini

$X \Rightarrow$ # games

Possible Values	Probability $P(X = x)$
4	
5	
6	
7	

1

probability distribution

$\Downarrow$

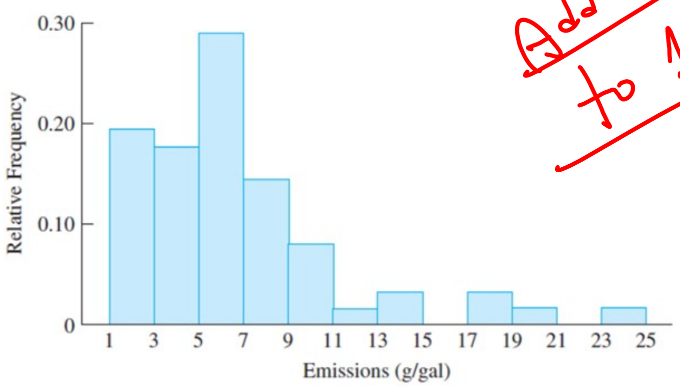
Discrete Random Variable

$\rightarrow$ Probability mass function pmf

Histogram

- A **histogram** is a graphic that gives an idea of the **shape** of a sample
- **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
- **Density** presents the **relative frequency divided by the class width**

Histogram



Add up to 1

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

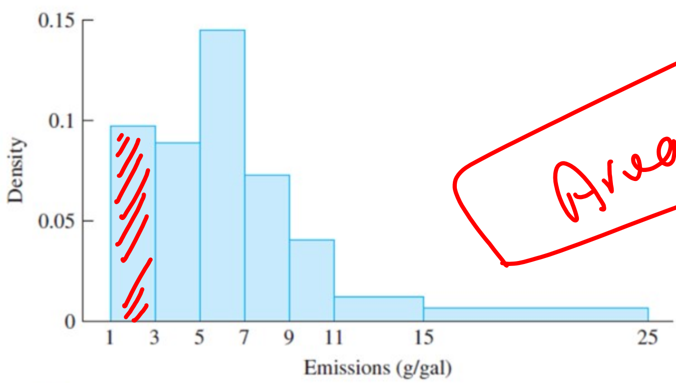
Histogram – unequal class width

$$\text{Density} = \frac{\text{Relative freq}}{\text{Class width}}$$

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



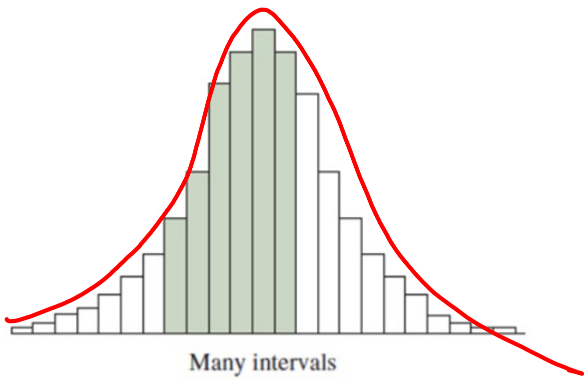
Area = 1

$$\text{Density} = \frac{\text{Relative freq}}{\text{class width}}$$

- 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**

Probability Density Function (PDF)

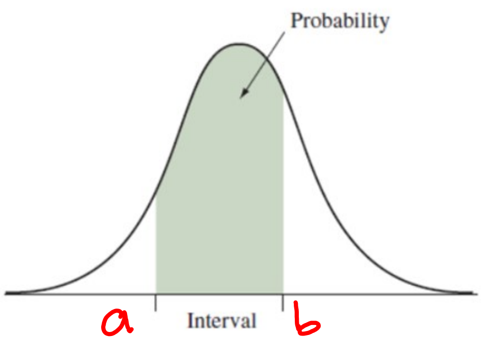
- As the **number of intervals increases**, with their **width narrowing**, the **shape** of the histogram **gradually approaches a smooth curve**.
- We'll use such curves to portray probability distributions of continuous random variables



Many intervals

Probability Density Function (PDF) → Continuous Random Variable

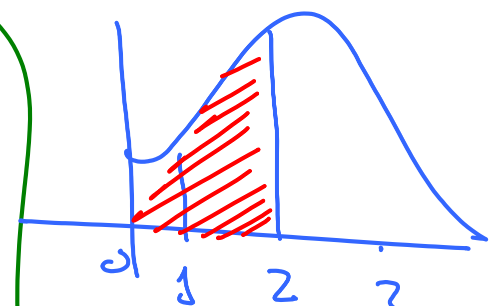
$$\int_{1.9}^{2.1} f(x)\,dx$$



Smooth curve approximation

- 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**

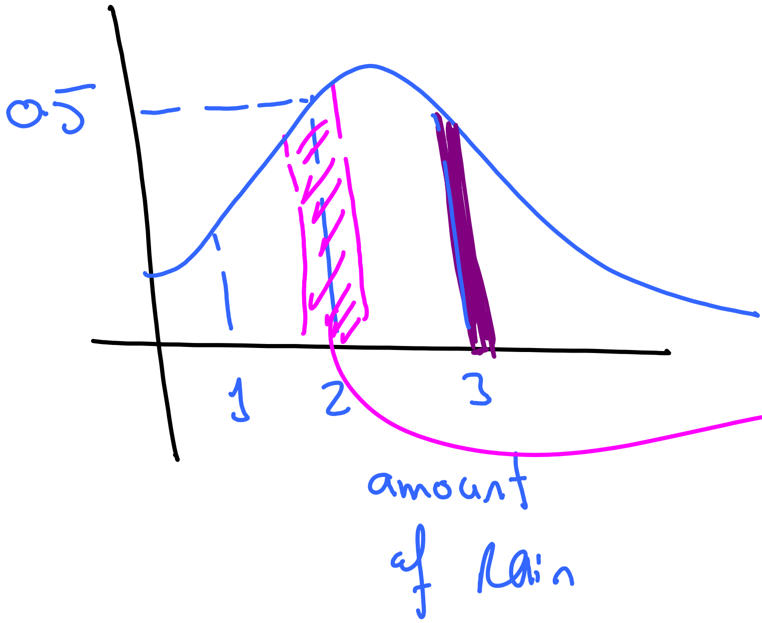
Y = Exact amount of Rain tomorrow



$P(Y<2)$

$\boxed{(a, b)}$

$$\int_a^b f(x)\,dx = \text{Area under Curve}$$



$P(|Y-2| < \underset{0.01}{0.1})$

$\Downarrow$

$P(\boxed{1.9 < Y < 2.1})$

$\int_{1.9}^{2.1} f(x)\,dx$

$\boxed{P(Y=3) = 0}$

$\int_a^b$

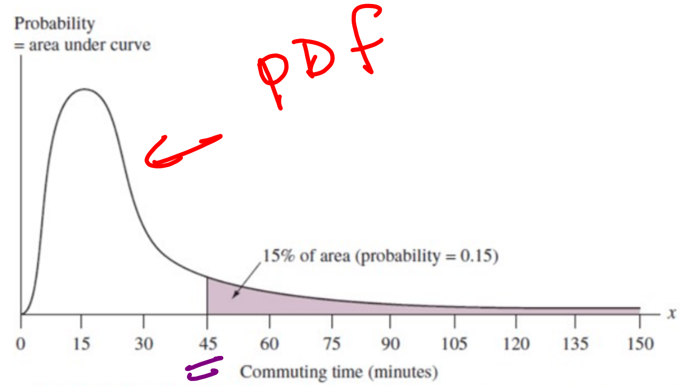
$P(Y=2) = \underline{\underline{0.5}}$ ✗

0 ✓

2.0000067
1.9995969

Example

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



pdf

▲ **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.

$P(X=45) = 0$

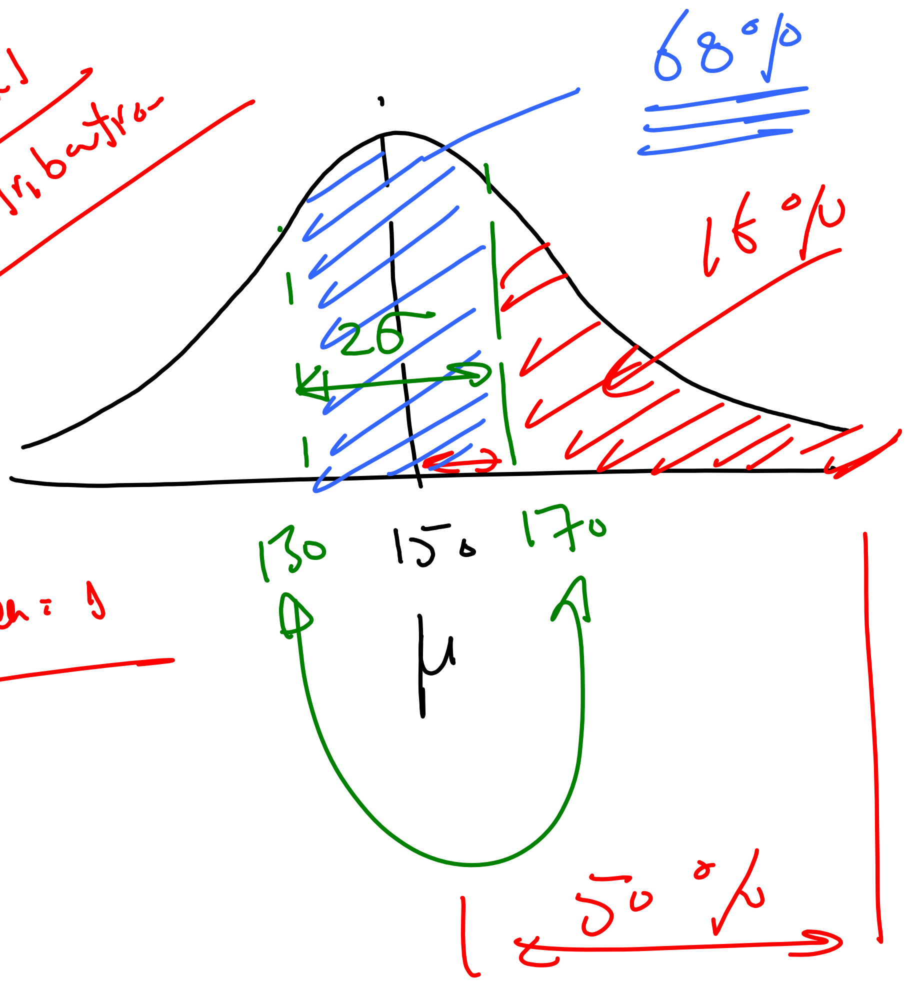
$P(X>45)$

$P(X\geq 45)$

$= 15\%$

normal distribution

Area = 1



$H \Rightarrow$ Random Variable height

$P(H > 170)$

$\mu = 150$

$\sigma = 20$

Z-table $\Rightarrow$ Coming

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