

Statistics

Lecture 17: | How to choose the endpoints of the class interval

Prof. Dr. Mostafa Elhosseini

<https://youtube.com/drmelhosseini>

| How to choose the endpoints of the class interval

- There is no **hard-and-fast rule** as to how to choose the endpoints of the class intervals
 - It is **good to have more intervals** rather than fewer, but
 - It is also **good to have large numbers of sample points** in the intervals
- Striking the proper balance is a matter of judgment and of **trial and error**

| How to choose the endpoints of the class interval

- When the **number of observations n is large** (several hundred or more), some have suggested that reasonable starting points for the number of classes may be **$\log_2 n$** (Binary Logarithm – Logarithm to base 2) or **$2 n^{1/3}$** .

- $\log_2 n = \frac{\ln n}{\ln 2}$

- When the **number of observations is smaller**, more classes than these are often needed

| Frequency – Relative Frequency

- **Frequency** represents the **numbers of data points** that fall into each of the **class intervals**.
- **Relative Frequency** represents the **frequencies divided by** the total **number of data points**,.
- The relative frequency of a class interval is the **proportion of data points that fall into the interval**

| Density

- **Density** represents the **relative frequency divided by the class width**.
- Note that when the **classes are of equal width**, the **frequencies**, **relative frequencies**, and **densities** are **proportional** to one another.
- When the **class intervals are of equal width**, the **heights** of the rectangles may be **set equal to the frequencies**, the **relative frequencies**, or the **densities**
- Since these three quantities are proportional, the **shape of the histogram will be the same** in each case

| Density

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-< 13	1	0.0161	0.0081
13-< 15	2	0.0323	0.0161
15-< 17	0	0.0000	0.0000
17-< 19	2	0.0323	0.0161
19-< 21	1	0.0161	0.0081
21-< 23	0	0.0000	0.0000
23-< 25	1	0.0161	0.0081

| Density

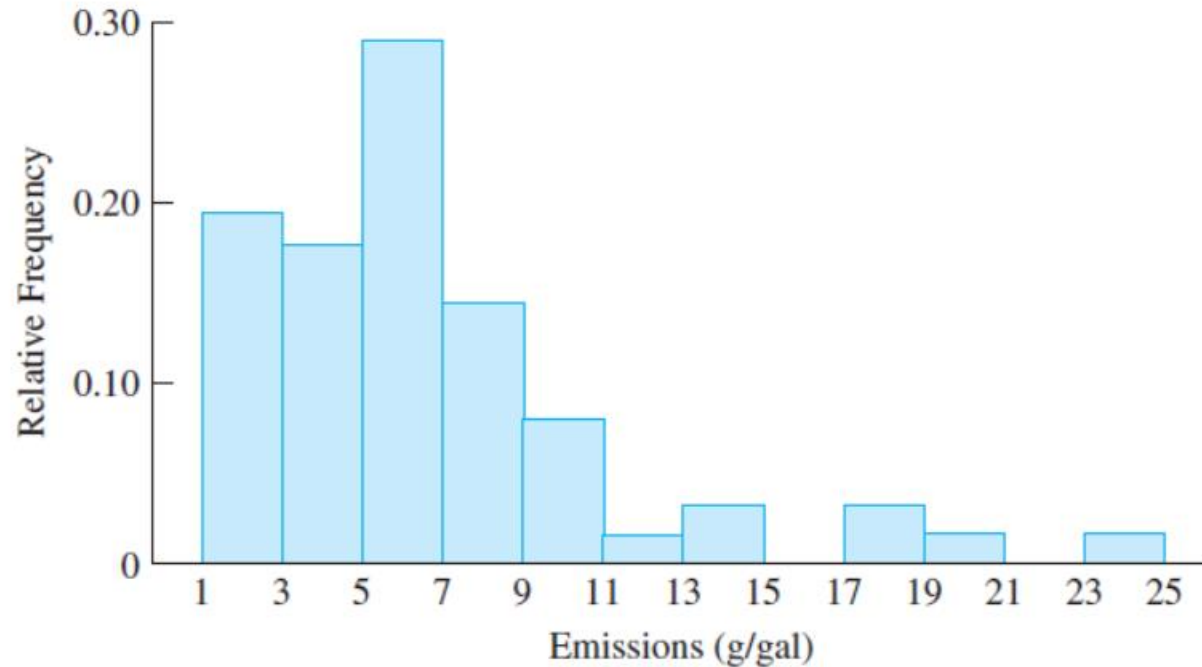
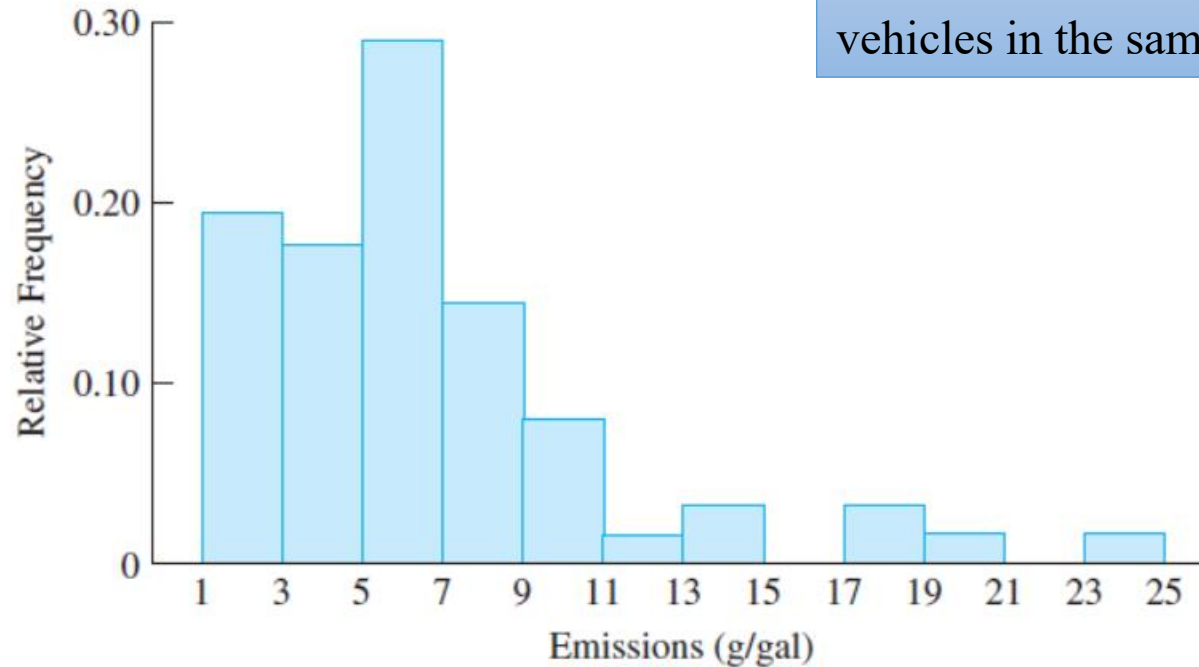


FIGURE 1.8 Histogram for the data in Table 1.4. In this histogram the heights of the rectangles are the relative frequencies. Since the class widths are all the same, the frequencies, relative frequencies, and densities are proportional to one another, so it would have been equally appropriate to set the heights equal to the frequencies or to the densities.

Example



Use the next histogram to determine the proportion of the vehicles in the sample with emissions between 7 and 11 g/gal

FIGURE 1.8 Histogram for the data in Table 1.4. In this histogram the heights of the rectangles are the relative frequencies. Since the class widths are all the same, the frequencies, relative frequencies, and densities are proportional to one another, so it would have been equally appropriate to set the heights equal to the frequencies or to the densities.

Example

Use the histogram in Figure 1.8 to determine the proportion of the vehicles in the sample with emissions between 7 and 11 g/gal.

Solution

The proportion is the sum of the relative frequencies of the classes spanning the range between 7 and 11. This is found by adding the heights of the rectangles for the two class intervals covered. The result is $0.1452 + 0.0806 = 0.2258$. Note that this result can also be obtained from the frequency table. The proportion of data points with values between 7 and 9 is 0.1452, and the proportion between 9 and 11 is 0.0806. The proportion between 7 and 11 is therefore equal to $0.1452 + 0.0806 = 0.2258$.

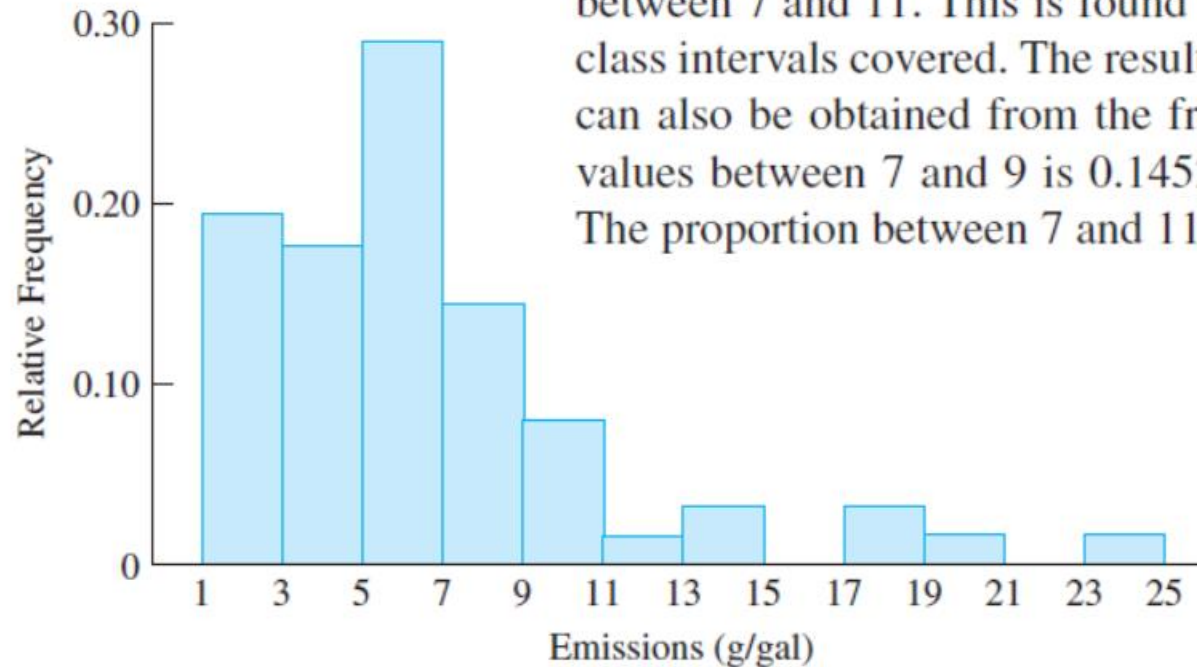


FIGURE 1.8 Histogram for the data in Table 1.4. In this histogram the heights of the rectangles are the relative frequencies. Since the class widths are all the same, the frequencies, relative frequencies, and densities are proportional to one another, so it would have been equally appropriate to set the heights equal to the frequencies or to the densities.