Density histogram*

Table 1 shows the frequency table of the monthly AIG stock price between January 2002 and December 2007. The data table is posted on Canvas (`table-3-1.txt`).

Stock price	**Count**	**Rel. freq. (%)**
\$ 45–\$ 50	2	2.8
\$ 50–\$ 55	3	4.2
\$ 55–\$ 60	13	18.1
\$ 60–\$ 65	16	22.2
\$ 65–\$ 70	24	33.3
\$ 70–\$ 75	13	18.1
\$ 75–\$ 80	1	1.4
Total	**72**	**100.1**

Table 1: Frequency table of the AIG monthly average stock price between January 2002 and December 2007. *Note.* Classes are left-closed, right-open. Relative frequencies do no add up to 100% because of rounding.

A histogram is a bar chart that displays the count of cases for each bin as the height of the bar over that bin (Sharpe et al., 2015, figure 3.1 p. 78). A relative frequency histogram displays the relative frequency for each bin as the height of the bar over that bin (Sharpe et al., 2015, figure 3.2 p. 79). When we start working with probability distributions it will be useful to work with a histogram that displays the relative frequency for each bin as the *area* of the bar over that bin. If the area of a bar represents relative frequency, how tall should the bar be? And what is the interpretation of the height of the bar?

The formula for the area of a rectangle is:

$$\text{area} = \text{base} \times \text{height}$$

As the area of each rectangular bar should represent relative frequency and the base is the width of each bin, we get:

$$\text{relative frequency} = \text{bin width} \times \text{height}$$

Solving for the height of the bar yields:

$$\text{height} = \frac{\text{relative frequency}}{\text{bin width}}$$

*Addendum to Sharpe et al. (2015, chapter 3)

The height of the bar in a density histogram is called **density**:

$$\text{density} = \frac{\text{relative frequency}}{\text{bin width}}$$

Let us compute density for the \$ 55–\$ 60 bin:

$$\text{density} = \frac{\text{relative frequency}}{\text{bin width}} = \frac{18.1\%}{\$\,5} = 3.62\% \text{ per } \$$$

In this case density has as units: percent per dollar. In general density has as units: percent per unit on the horizontal axis. Now what does the statement "density of the \$ 55– \$ 60 bin is 3.62 percent per dollar" mean? The area of the bar over the \$ 55–\$ 60 bin is 18.1 percent. Slice up the bar into five vertical strips, each one dollar and 3.62% per dollar tall. Assume that the 13 cases that fall into the \$ 55–\$ 60 bin are uniformly distributed over the bin. The area of each vertical strip of one dollar wide is $\frac{1}{5}$ of 18.1 percent, that is, 3.62 percent. That also follows from applying the formula for the area of a rectangle to each vertical strip of 1 dollar wide:

$$\text{area} = \text{base} \times \text{height} = \$\,1 \times 3.62\% \text{ per } \$ = 3.62\%$$

The interpretation of a density of 3.62 percent per dollar is as follows: Between \$ 55 and \$ 60 each interval of 1 dollar wide contains approximately 3.62 percent of the values. The interval \$ 55–\$ 56 contains about 3.62 percent of the values, and the same is true for the intervals \$ 56–\$ 57, \$ 57–\$ 58, \$ 58–\$ 59, and \$ 59–\$ 60.

Statisticians usually express density not as a percentage per unit on the horizontal axis but as a *decimal proportion* per unit on the horizontal axis:

$$3.62\% \text{ per } \$ = \frac{3.62}{100} \text{ per } \$ = 0.0362 \text{ per } \$$$

Verify the values of the densities for the other bins. Table 2 shows all densities.

Stock price	**Count**	**Rel. freq. (%)**	**Density (dec. prop. per \$)**
\$ 45–\$ 50	2	2.8	0.0056
\$ 50–\$ 55	3	4.2	0.0084
\$ 55–\$ 60	13	18.1	0.0362
\$ 60–\$ 65	16	22.2	0.0444
\$ 65–\$ 70	24	33.3	0.0667
\$ 70–\$ 75	13	18.1	0.0362
\$ 75–\$ 80	1	1.4	0.0028
Total	**72**	**100.1**	—

Table 2: Frequency table (including densities) of the AIG monthly average stock price between January 2002 and December 2007. *Note.* Classes are left-closed, right-open. Relative frequencies do no add up to 100% because of rounding.

To generate a density histogram in R add the option `freq=FALSE` to the `hist()` function:

```
hist(x, freq=FALSE)
```

To obtain the values of the densities add the option `plot=FALSE`:

```
hist(x, plot=FALSE)
```

Figure 1 shows the density histogram. Note that the shape is identical to the shape of the histogram that shows the counts (Sharpe et al., 2015, figure 3.1 p. 78) and of the relative frequency histogram (Sharpe et al., 2015, figure 3.2 p. 79). However, a density histogram has as a property that the area of the histogram over all the values of the variable is equal to 1.

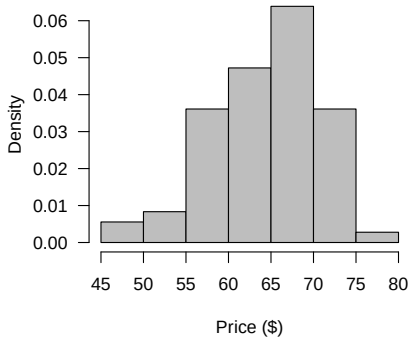

Figure 1: Density histogram of the AIG monthly average stock price between January 2002 and December 2007

References

Sharpe, N. R., De Veaux, R., and Velleman, P. (2015). *Business Statistics*. Pearson Education, 3rd edition.