Bookmark this page

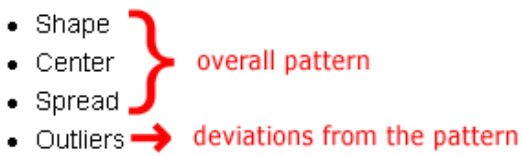
Histogram: Shape

Learning Objective: Generate and interpret several different graphical displays of the distribution of a quantitative variable (histogram, stemplot, boxplot).

Learning Objective: Summarize and describe the distribution of a quantitative variable in context: a) describe the overall pattern, b) describe striking deviations from the pattern.

Interpreting the Histogram

Once the distribution has been displayed graphically, we can describe the overall pattern of the distribution and mention any striking deviations from that pattern. More specifically, we should consider the following features of the distribution:



We will get a sense of the overall pattern of the data from the histogram's center, spread and shape, while outliers will highlight deviations from that pattern.

Shape

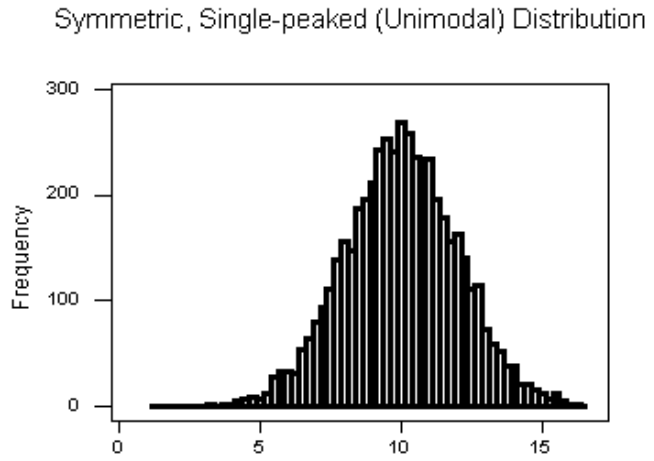
When describing the shape of a distribution, we should consider:

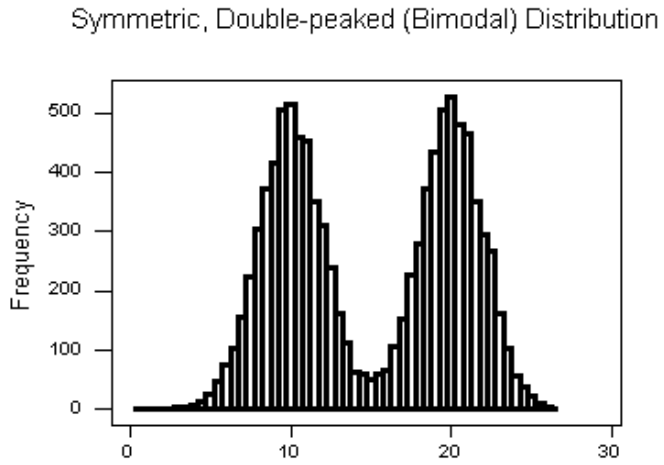
1. **Symmetry/skewness** of the distribution.

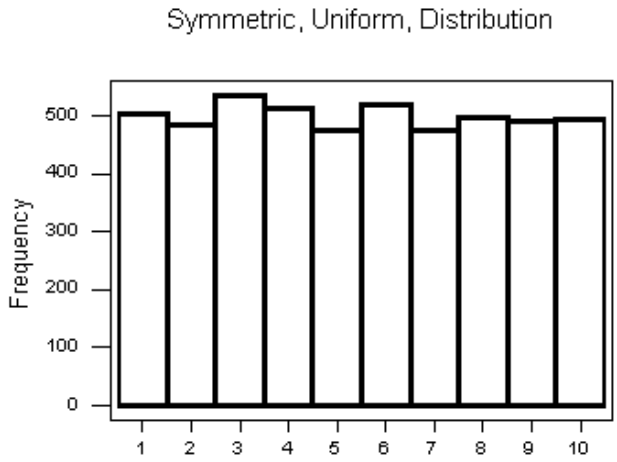
2. **Peakedness (modality)**—the number of peaks (modes) the distribution has.

We distinguish between:

Symmetric Distributions



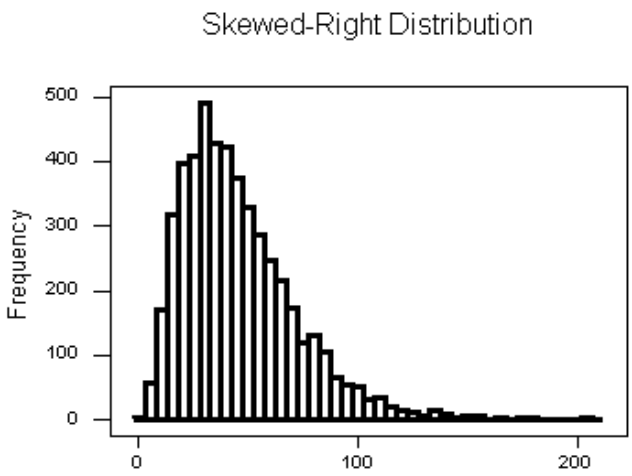




Note that all three distributions are symmetric, but are different in their modality (peakedness). The first distribution is **unimodal**—it has one mode (roughly at 10) around which the observations are concentrated. The second distribution is **bimodal**—it has two modes (roughly at 10 and 20) around

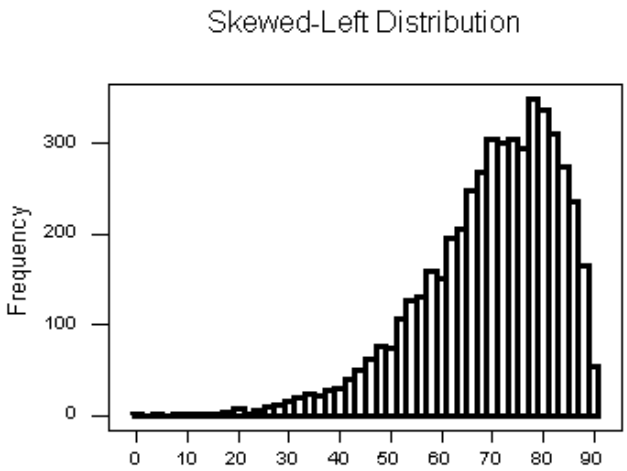
which the observations are concentrated. The third distribution is kind of flat, or **uniform**. The distribution has no modes, or no value around which the observations are concentrated. Rather, we see that the observations are roughly uniformly distributed among the different values.

Skewed Right Distributions



A distribution is called **skewed right** if, as in the histogram above, the right tail (larger values) is much longer than the left tail (small values). Note that in a skewed right distribution, the bulk of the observations are small/medium, with a few observations that are much larger than the rest. An example of a real-life variable that has a skewed right distribution is salary. Most people earn in the low/medium range of salaries, with a few exceptions (CEOs, professional athletes etc.) that are distributed along a large range (long "tail") of higher values.

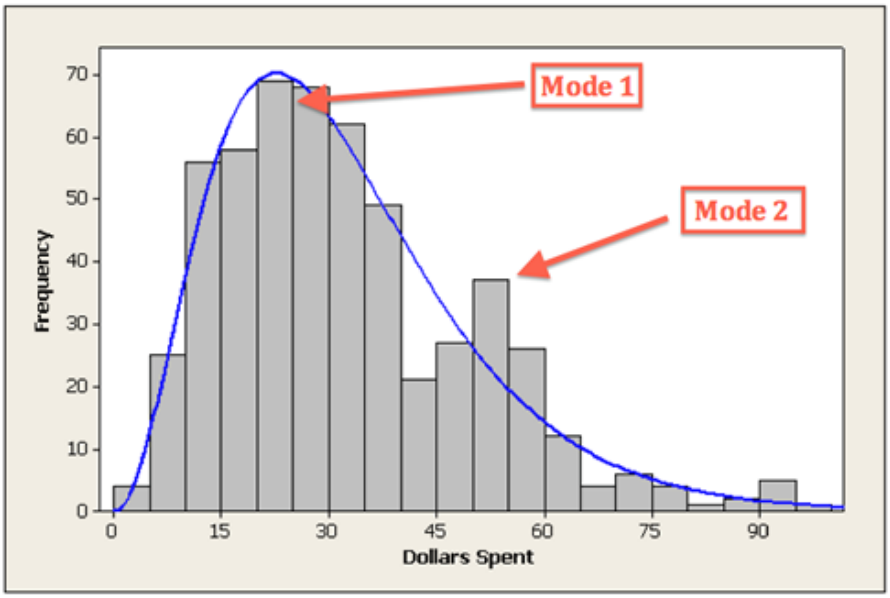
Skewed Left Distributions



A distribution is called **skewed left** if, as in the histogram above, the left tail (smaller values) is much longer than the right tail (larger values). Note that in a skewed left distribution, the bulk of the observations are medium/large, with a few observations that are much smaller than the rest. An example of a real life variable that has a skewed left distribution is age of death from natural causes (heart disease, cancer etc.). Most such deaths happen at older ages, with fewer cases happening at younger ages.

Comments:

1. Note that skewed distributions can also be bimodal. Here is an example. A medium size neighborhood 24-hour convenience store collected data from 537 customers on the amount of money spend in a single visit to the store. The following histogram displays the data.



   Note that the overall shape of the distribution is skewed to the right with a clear mode around $25. In addition it has another (smaller) “peak” (mode) around $50-55. The majority of the customers spend around $25 but there is a cluster of customers who enter the store and spend around $50-55.

2. If a distribution has more than two modes, we say that the distribution is **multimodal**.

Recall our grades example below. As you can see from the histogram, the grades distribution is roughly symmetric.