

2-2 Frequency distributions

Def: A frequency **distribution** (or frequency table) lists classes (or categories) of values, along with frequencies (or counts) of the number of values that fall into each class.

A data is considered **discrete** if it is finite or a countable number.

A data is considered **continuous** if is an uncountable number, represented by an interval on the number line.

Example 1:

You conduct a study of the number of calls a sales person makes in one day. The possible values are 0, 1, 2, 3, 4, and so on. (discrete data)

You conduct a study to measure the time (in hours) a salesperson spends making calls in one day. Because the time spent making sales calls can be any number from 0 to 24 (including fractions and decimals), this data is a continuous data. You can represent its values with an interval on a number line, but you cannot list all the possible values.

- If the data are discrete and have few different values, the categories of data will be the observations.
- If the data are discrete, but there are many different values of the variables, categories of data (called classes) must be created using intervals of numbers.
- If the data are continuous, the categories of data must be created using intervals of numbers.

Constructing a Frequency Distribution from a Data Set:

Guidelines:

1. Decide on the number of classes to include in the frequency distribution. The number of classes should be between five and twenty.

2. Find the **class width**. The class width is the difference between the maximum and minimum data entries, divided by the number of classes, and rounded up to the next convenient number.
3. Find the **class limits**. A lower class limit is the least number that can belong to a specific class and an upper class limit is the greatest. Use the minimum data entry as the lower limit of the first class. To find the remaining lower limits, add the class width to the lower limit of the preceding class.
4. Use tally marks to sort the data entries into classes.
5. Count each of the tally marks to find the total frequency, f , for each class.

Example 2:

The following sample data set list the number of minutes 25 people spent reading the newspaper in a day. Construct a frequency distribution that has 5 classes.

7	39	13	9	25	8	22	0	2	18	2	30	7
35	12	15	8	6	5	29	0	11	39	16	15	

Definition:

The **midpoint** of a class is the sum of the lower and upper limits of the class divided by two.

The **relative frequency** of a class is the portion or percent of that data that falls in that class. To find the relative frequency of a class, divide the frequency f by the sample size (sum of all frequencies), n .

The **cumulative frequency** of a class is the sum of the frequency for that class and all previous classes. The cumulative frequency of the last class is equal to the sample size n .

Example 3: Using the frequency distribution constructed in Example 2, find the midpoint, relative frequency, and cumulative frequency for each class.

Organizing continuous data into a frequency and relative frequency distribution:

We use **class boundaries** for continuous data to separate classes. Class boundaries are obtained as follows:

Find the size of the gap between the upper class limit of one class and the lower class limit of the next class. Add half of that amount to each upper class limit to find the upper class boundaries; subtract half of that amount from each lower class limit to find the lower class boundaries.

Example 4: Using the frequency distribution constructed in example 3, find the class boundaries.

2-3 Histograms

To visualize data we graph the pictures of distributions. Among the different types of graphs in this chapter, we start with histogram.

Def: A **histogram** is a graph that uses bars to portray the frequencies or the relative frequencies of the possible outcomes for the numerical data. In which the horizontal scale represents classes and the vertical scale represents frequencies. The heights of the bars correspond to the frequency values, and the bars are drawn adjacent to each other (without gaps).

Because consecutive bars of a histogram must touch, bars must begin and end at class boundaries instead of class limits.

Example 5: Draw a frequency histogram for the frequency distribution in example 2.

Relative Frequency Histogram

A **relative frequency histogram** has the same shape and horizontal scale as a histogram, but the vertical scale is marked with relative frequencies instead of actual frequencies.

Example 6: Draw a relative frequency histogram for the frequency distribution in example 3.

2-4 Stem-and-leaf plot

A stem-and-leaf plot is another way to represent numerical data graphically.

We use the following steps to construct a stem-and-leaf plot.

1. The stem of the graph will consist of the digits to the left of the right most digit. The leaf of the graph will be the right most digit. (sometimes it is necessary to modify the method of choosing the stem if a different class width is desired)
2. Write the stems in a vertical column in increasing order. Draw a vertical line to the right of the stems.
3. Write each leaf corresponding to the stem to the right of the vertical line.
4. Write leaves in ascending order.

Example 7: consider the following test grades. Construct a stem-and-leaf plot.

67 72 85 75 89 89 88 90 99 100

By turning the page on its side, we can see a distribution of these data. The greatest advantage of the stem-and-leaf plot is that we can see the distribution of data and yet keep all information in the original list.