Sullivan Section 2.2

Gene Quinn

Organizing Quantitative Data

quantitative data consists of *numerical* measures that characterize individuals.

Organizing Quantitative Data

quantitative data consists of *numerical* measures that characterize individuals.

By *numerical* measures, we mean measures for which arithmetic operations make sense.

Organizing Quantitative Data

quantitative data consists of *numerical* measures that characterize individuals.

By *numerical* measures, we mean measures for which arithmetic operations make sense.

So a measure like arrivals at a collection of stores would be considered numerical because adding the number of arrivals at two stores would still give a meaningful number.

Discrete and Continuous Data

quantitative data can be grouped into two major categories:

- discrete
- continuous

Discrete and Continuous Data

quantitative data can be grouped into two major categories:

- discrete
- continuous

Discrete data typically assume only a limited number of values. These values can be treated as categories.

Discrete and Continuous Data

quantitative data can be grouped into two major categories:

- discrete
- continuous

Discrete data typically assume only a limited number of values. These values can be treated as categories.

Continuous variables can potentially assume a very large number of values or even, in principle at least, an infinite number of values on some interval.

Continuous or Discrete?

With quantitative data, one of the first steps is deciding whether the data is continuous or discrete. Meaures like household income, weight, credit card debt, etc. would be examples of continuous data.

Continuous or Discrete?

With quantitative data, one of the first steps is deciding whether the data is continuous or discrete. Meaures like household income, weight, credit card debt, etc. would be examples of continuous data.

One indication that the data is continuous might be that few values repeat. A continuous variable would generally not be a good candidate for the horizontal axis in a histogram, unless it was grouped into ranges.

Continuous or Discrete?

With quantitative data, one of the first steps is deciding whether the data is continuous or discrete. Meaures like household income, weight, credit card debt, etc. would be examples of continuous data.

One indication that the data is continuous might be that few values repeat. A continuous variable would generally not be a good candidate for the horizontal axis in a histogram, unless it was grouped into ranges.

With discrete data, on the other hand, there is usually considerable reptition in the data values. Discrete data would fit well as the horizontal axis in a histogram as it is.

Table Summaries for Discrete Data

Like nonnumeric data, for discrete numeric data we can construct a frequency table, or if we prefer a relative frequency table.

Table Summaries for Discrete Data

Like nonnumeric data, for discrete numeric data we can construct a frequency table, or if we prefer a relative frequency table.

The values of the variable that occur in the data form the classification groups or row labels for the table.

Histograms for Discrete Data

A **histogram** is a kind of chart constructed by drawing a rectangle for each class of data.

The textitheight of the rectangle represents the frequency or relative frequency of one value in the sample.

A histogram differs from a bar chart in that the rectangles should touch each other and should each be the same width.

The histogram should approximate a curved shape using rectangles.

Table Summaries for Continuous Data

For continuous numeric data, we usually have to collect values into intervals called **classes**.

Table Summaries for Continuous Data

For continuous numeric data, we usually have to collect values into intervals called **classes**.

The smallest possible value within a given class is called the **lower class limit**.

Table Summaries for Continuous Data

For continuous numeric data, we usually have to collect values into intervals called **classes**.

The smallest possible value within a given class is called the **lower class limit**.

The largest possible value within a class is called the **upper class limit**.

A class with no limit to the size of possible values is called **open ended**.

Table Summaries for Continuous Data

For continuous numeric data, we usually have to collect values into intervals called **classes**.

The smallest possible value within a given class is called the **lower class limit**.

The largest possible value within a class is called the **upper class limit**.

A class with no limit to the size of possible values is called **open ended**.

The difference between consecutive lower class limits is called the **class width**.

Table Summaries for Continuous Data

For continuous data, once the classes are defined, we count the number of values that fall in each class. **classes**.

Table Summaries for Continuous Data

For continuous data, once the classes are defined, we count the number of values that fall in each class. **classes**.

The row identifier in the table will usually be the lower and upper class limits for the class.

Table Summaries for Continuous Data

For continuous data, once the classes are defined, we count the number of values that fall in each class. **classes**.

The row identifier in the table will usually be the lower and upper class limits for the class.

For a frequency table, the number in the column will be the number of values that fall in this class.

Stem and Leaf Plots

A **stem and leaf** plot is a type of graphical representation of numeric data. **classes**.

Stem and Leaf Plots

A **stem and leaf** plot is a type of graphical representation of numeric data. **classes**.

A stem and leaf plot tends to reveal the shape of the distribution well.

Stem and Leaf Plots

A **stem and leaf** plot is a type of graphical representation of numeric data. **classes**.

A stem and leaf plot tends to reveal the shape of the distribution well.

Unlike histograms, the original data can be reconstructed from the stem and leaf plot.

Distribution Shapes

We will consider four major distribution shapes as revealed by histograms or possibly stem and leaf plots:

- uniform (all rectangles nearly the same height)
- bell-shaped (tops of the rectangles form a bell-shaped curve)
- skewed right (like bell-shaped, but with a **tail** stretching out to the right)
- skewed left (like bell-shaped but with a **tail** stretching out to the left.

Time Series Plots

A **time series plot** is constructed by plotting a point directly above the point in time on the horizontal axis when the variable was measured, and at the value of the measurement on the vertical axis.

Once plotted, successive points are joined by lines.