Section 2.2

Objectives

By the end of this section, you will be able to...

1. organize quantitative data into tables
2. construct histograms for discrete and continuous data
3. draw stem-and-leaf plots
4. draw dot plots
5. identify the shape of a distribution

Like qualitative data in the last section, quantitative data can (and should) be organized into tables. We'll break this page up into two parts - discrete and continuous.

Organizing Discrete Data into Tables

If you recall from Section 1.2,

> A **discrete variable** is a quantitative variable that has either a finite number of possible values or a countable number of values. (*Countable* means that the values result from counting - 0, 1, 2, 3, ...)

Since we can list all the possible values (that's essentially what *countable* means), one way to make a table is just to list the values along with their corresponding frequency.

Example 1

Here's some data I collected from a previous students Mth120 course. It refers to the number of children in their family (including themselves).

2	2	2	4	5	3	3	3	3
2	1	2	3	5	3	4	3	1
2	3	5	3	2	1	3	2	

An easy way to compile the data would then be to make a frequency or relative frequency table as we did before.

children	frequency	relative frequency
1	3	3/26 ≈ 0.12
2	8	8/26 ≈ 0.31
3	10	10/26 ≈ 0.38
4	2	2/26 ≈ 0.08
5	3	3/26 ≈ 0.12

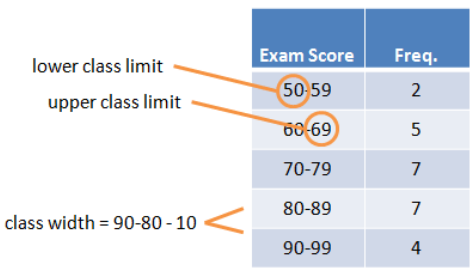
Example 2

A good example might be the scores on an exam, ranging from 1-100. Here are some data from a past Mth120 class.

62	87	67	58	95	94	91	69	52
76	82	85	91	60	77	72	83	79
63	88	79	88	70	75	87		

In this case, we'll have to set up intervals of numbers called **classes**. Each class has a **lower class limit** and an **upper class limit**, along with a **class width**. The class width is the difference between successive lower class limits.

To be consistent, the class width should be same for each class. One good option might look something like this:



Exam Score	Freq.
50-59	2
60-69	5
70-79	7
80-89	7
90-99	4

Organizing Continuous Data into Tables

Organizing continuous data is similar to organizing multi-valued discrete data. We have to form classes which don't overlap. I usually try to design a class width that's either logical (i.e. 10 points for grades above) or so that I have 5-8 classes when complete.

Example 3

For this example, let's consider the average commute for each of the 50 states. The data below show the average daily commute of a random sample of 15 states.

23.1	18.3	23.2	19.9	26.6
24.8	23.1	23.2	22.7	29.4
22.3	30.0	25.8	21.9	16.7

Source: US Census

Do you know why this is a continuous random variable and not discrete? (Hint: It's *not* because of the decimal.)

I think I know!

To make a frequency or relative frequency for continuous data, we use the same strategy we'd use for multi-valued discrete data.

average commute	frequency	relative frequency
16-17.9	1	1/15 ≈ 0.07
18-19.9	2	2/15 ≈ 0.13
20-21.9	1	1/15 ≈ 0.07
22-23.9	6	6/15 = 0.40
24-25.9	2	2/15 ≈ 0.13
26-27.9	1	1/15 ≈ 0.07
28-29.9	1	1/15 ≈ 0.07
30-31.9	1	1/15 ≈ 0.07

Technology

Here's a quick overview of how to create frequency and relative frequency tables for quantitative data in StatCrunch.

Discrete Data

1. Enter or import the data.
2. Select **Stat** > **Tables** > **Frequency**.
3. Select the column(s) you want to summarize and click **Next**.
4. Add any modifications for an "Other" category and how to order the categories, and click **Calculate**.

Continuous or Multi-valued Discrete Data:

1. Enter or import the data.
2. Select **Data** > **Bin Column**.
3. Select the column containing the data, select "Use fixed width bins", and set the lowest class limit (*Start bins at:*) and class (*bin*) width.
4. Click **Calculate.**
5. Select **Stat** > **Tables** > **Frequency**.
6. Select the newly created bin column and click **Calculate**.*

* Note that these classes *seem* to overlap, but that the class "0-k" does not include *Mk*.

Single-valued Histograms

To display quantitative data, we need a new type of chart, called a **histogram**. Histograms look similar to bar graphs, but they have some distinct differences - and for good reason.

> A **histogram** is constructed by drawing rectangles for each class of data. The height of each rectangle is the frequency or relative frequency of the class. *The width of each rectangle is the same and the rectangles touch each other.*

The rectangles need to touch in a histogram because we want to imply that the classes are adjacent. In a bar graph, a favorite color of "blue" isn't really adjacent to "red", even though we might put it that way in a bar graph. For quantitative data like the data used in Example 1 earlier this section, the value 2 really is next to the value 3.

Let's take a closer look at that example.

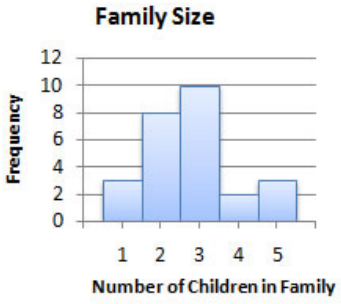
Example 4

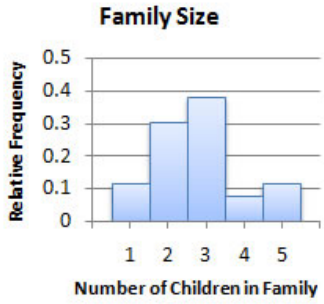
children	frequency	relative frequency
1	3	3/26 ≈ 0.12
2	8	8/26 ≈ 0.31
3	10	10/26 ≈ 0.38
4	2	2/26 ≈ 0.08
5	3	3/26 ≈ 0.12

To make a histogram, we make what looks like a bar graph with a couple key differences:

1. rectangles must touch
2. class labels are underneath the rectangle

Here's what they'd look like for our example data:





Technology

Here's a quick overview of how to create histograms for single-valued discrete data using StatCrunch.

1. Enter or import the data.
2. Select **Graphics** > **Histogram**
3. Select the column(s) you want to summarize and click **Next**.
4. Set the *Type,* lower class limit (*Start bins at:*).
5. Set the class width (*bin*) to 1, and click **Calculate**.