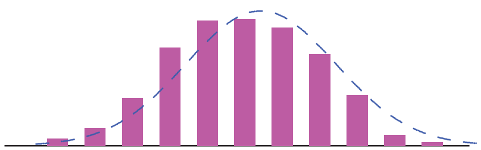
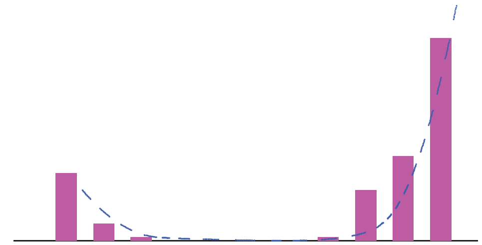
6. Normal Tables

Sometimes the graph of data will appear to fall into a regular pattern. There are certain kinds of patterns that occur over and over again in data. One of the most important of these recurring patterns is called the "normal distribution" or "bell-shaped curve."



Examples of data that might follow a normal distribution are height, weight, income, and IQ scores. Many standardized measurements turn out to be normally distributed.

However, just because this distribution *often* occurs does not mean it *always* occurs. For example, if there were a "handedness" scale it might look like



reflecting the fact that almost no one is truly ambidextrous: about 7% of the population is left-handed and about 93% is right-handed. Interestingly, "pawed-ness" in mice is normally distributed. (How do you think you might test for this in mice?)

Properties of normal curves:

- Normal curves are symmetric about the mean μ.
- About 95% of the observations fall between

$$\mu - 2\sigma \quad \text{and} \quad \mu + 2\sigma$$

- About 68% of the observations fall between

$$\mu - \sigma \quad \text{and} \quad \mu + \sigma$$

- The curves are completely characterized by the mean μ and the standard deviation σ.

The statement “the curves are completely characterized by the mean and standard deviation” means that the *formula* for the normal curve involves only these two parameters:

$$\text{normal curve} = \frac{1}{\sqrt{2\pi}\sigma} e^{\left(\frac{x-\mu}{\sigma}\right)^2}.$$

Normal curves were first understood as an "error function" – as the natural variation that occurs in all physical measurements. The German mathematician Karl Gauss was the first to see the normal curve as something that could be graphed. Because the German word for "numbers" is "zahlen" the values associated with the normal curve are sometimes call "*z-values*."

Prior to the introduction of the euro, the most common German banknote was the 10 deutsch mark note which included both Gauss' face and the above formula.

The *area under the curve* corresponds to population percentiles. For example, 95% of the *area* under the curve is between

$$\mu - 2\sigma \quad \text{and} \quad \mu + 2\sigma$$

and hence 95% of the observations in a normal population also fall in this range. There are numerical techniques for finding the area under curves – this is a major topic in calculus. Using these techniques it is possible to compute the percentile corresponding to an observation from a normally distributed population. Instead of doing this computation for each possible observation, statisticians instead build a table of possible observations and corresponding percentiles.

There are two categories of problems you will learn how to solve using normal tables: “outside-in" and “inside-out" problems. The reason for the words “outside-in" and “inside-out" has to do with how one uses the tables and will be clearer after we do some problems.

Finding percentiles (outside-in problems). In these you will be given:
- the mean
- the standard deviation
- a “raw” score (a measurement on a member of the population)

And you will be looking for
- the percentile which corresponds to the given measurement.

Finding measurements (inside-out problems.)
In these problems you will be given:
- the mean
- the standard deviation
- a percentile

And you will be looking for
- the “raw score” or measurement which corresponds to the given percentile.

These are the two classes of problems which you will learn how to solve in this segment.
As we have previously seen, percentiles by themselves are not particularly useful objects. However, the *techniques* will turn out to be fundamental in working other kinds of problems.