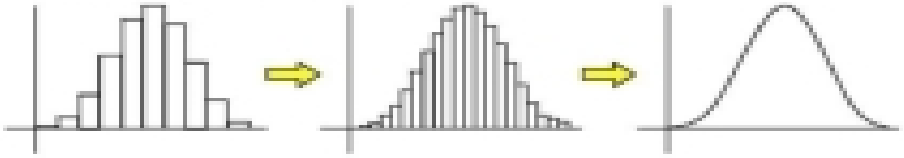
Probability Density Curves

Recall that a continuous statistical variable is a variable that can take on any value within a certain range of numbers to however many decimal places you can measure. To make a histogram graph out of the data for a continuous variable, we must divide the data into classes.

The histogram can give us an idea of how likely it is that a randomly chosen response from a survey will fall into one of the classes: the higher the bar for a particular class, the more likely that result is. We can also use a histogram to tell us how likely we are to choose a response that falls within a given range of classes. We do this by considering the heights of all the bars within that range — the sum of these bar heights tells you the chances of a random response falling within the range of interest. More properly, the area taken up by the bars gives us the probability, since the sum of the areas of all bars represents 100% of the data.

The more data you have, the more detailed you can make your histogram. You can make the classes narrower and the bars that result will still have meaningful heights. What if you had an infinite amount of data, and the data came from a regular, consistent source? What we mean by regular and consistent is that all the data has a similar centre (mean, μ) and a similar spread (standard deviation, σ). We could make narrower and narrower classes until they had no width at all, and the jagged corners at the top of the histogram would become smooth curves:



A curve like this is called a **density curve**. It is an idealized mathematical model of the behaviour of data. **The total area under any density curve is 1** because it represents all the possible data values that the variable can take (the sum of all possible outcomes of an event must equal 1). Also, a density curve will always be above the horizontal axis since having any part of it be below the axis would imply a negative probability of a value occurring, which is meaningless.

We can use a density curve to find the probability that a randomly chosen representative of a data set has a value higher or lower than a specified number, or between specified numbers. **To find a probability using a density curve, we make a vertical slice through the curve** at any value at the edge of an interval we're interested in, **and we calculate the area under the curve** within the interval. This is an extremely important concept to master, because the entire rest of the stats course will

Authored by Darren Rigby