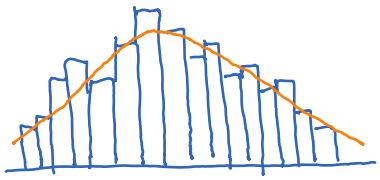
Lecture 22

Wednesday, November 6, 2024 1:32 PM

Recall that the histogram encodes all the information of a data set.

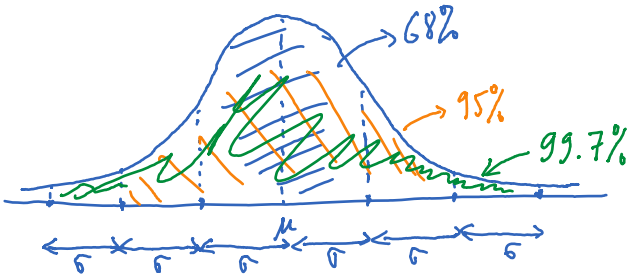


A distribution curve is a curve that fits the histogram. A *normal* distribution curve is the graph of a function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Here, μ is the center of the curve and σ is the standard deviation of the curve.
Note that not all bell-shaped curves are normal distribution curves. Only those that are a graph of the function above are normal distribution curves.

σ can be interpreted from the curve as follows: about 68% of all data lie within one standard deviation from the mean μ. So, about 68% of all data lie between $\mu - \sigma$ and $\mu + \sigma$. About 95% of all data lie within two standard deviations from the mean. So, about 95% of all data lie between $\mu - 2\sigma$ and $\mu + 2\sigma$. About 99.7% of all data lie within two standard deviations from the mean. So, about 99.7% of all data lie between $\mu - 3\sigma$ and $\mu + 3\sigma$.



Example:
SAT scores curved to fit a normal distribution curve with $\mu = 500$ and $\sigma = 100$. If you get 600 points, what is your percentile? In other words, what is the percentage of all students who score the same as you or less?

There are 50% of students who score less than 500.
There are about 68% of students who score between 400 and 600.
Due to symmetry, there are about 34% of students who score between 500 and 600.
Therefore, there are about 50%+34%=84% of students who score 600 or less.

Note that you can use Desmos or GeoGebra to sketch the distribution curve that ETS uses to curve the SAT score. Simply graph the function f above using $\mu = 500$ and $\sigma = 100$.