6.1 Introducing Normally Distributed Variables

Up to now, we've dealt with *discrete random variables*, variables that take on only a finite (or countably infinite—we didn't do these) number of values. "how many?"

A *continuous random variable* takes on all real values in an interval. "how much?"

Examples: Continuous random variables:
Heights of people, lifetimes of light bulbs, lengths of phone calls.

Discrete random variables:
Number of defective clocks in a sample, number of students who say they like math, houses in a neighborhood.

For discrete random variables, the probability assignments for each possible value of the variable are listed in a table or computed using a formula.

For continuous random variables, we compute the probability of the variable falling within a certain interval, instead of the probability of individual values. The probability is given by the area under the curve of the *probability density function* for that distribution.

Suppose we conduct the experiment a large number of times, record the values of the variable, and use a histogram to summarize the frequencies of the values obtained. If we draw a smooth curve that connects the tops of all the histogram bars, this curve would approximate the probability density function for that experiment.

Probability Density Functions (for Continuous Random Variables):

For an equation to be a probability density function, it must satisfy the following properties:

1) Total area under the graph of the equation = 1

2) The graph of the equation must never dip below the *x*-axis (i.e., all y-values must be greater than or equal to 0.)

These are analogous to the requirements that the probabilities in a finite discrete distribution must have a sum of 1, and that all probabilities must be between 0 and 1.

The area under the graph (between the graph and the *x*-axis) over an interval represents the probability of observing a value of the random variable in that interval.

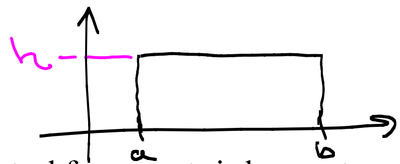
The uniform distribution:

In the uniform distribution, all values in an interval are equally likely. In a histogram from a uniform distribution, all bars should be approximately the same height.

Suppose a continuous variable is equally likely to take on any value between *a* and *b*. Because the uniform distribution is rectangular in shape, and because the total area under the curve must be 1, then the height of the rectangle must be $\frac{1}{b-a}$.

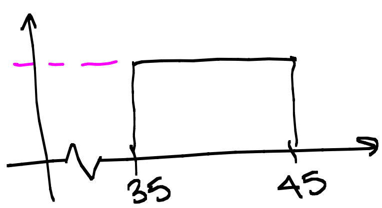
(width) (height) = 1

width = b − a, so (b − a)(height) = 1 ⟹ height = $\frac{1}{b-a}$



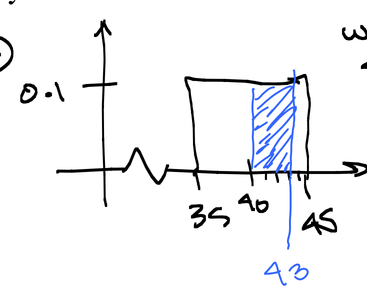
Example 1: Suppose the thickness of the coating on steel fence posts is known to range from 35 microns to 45 microns. Suppose the coating thickness follows a uniform distribution.

a) What is the probability that a randomly selected steel fence post will have a coating between 40 and 43 microns in thickness?
b) What is the probability the post will have a coating less than 40 microns in thickness?
c) What is the probability the post will have a coating exactly 40 microns in thickness?



Area = 1
width = 45 − 35 = 10
h (10) = 1
h = $\frac{1}{10}$ = 0.1

(a)



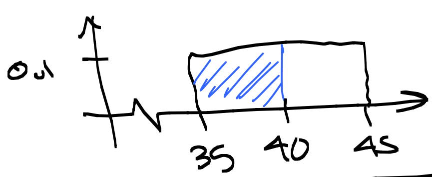
width of shaded area = 43 − 40 = 3

Probability = Shaded Area = 3(0.1) = 0.3

X = thickness,

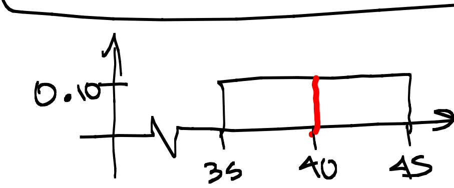
$\boxed{P(40 < X < 43) = 0.3}$

(b)



$P(X < 40) = \frac{1}{2} = \boxed{0.5}$

(c)



so technically, according to this model, $P(X = 40) = \boxed{0}$

The normal distribution:

For many continuous random variables, most of the data points fall near the center, while fewer data points fall near the ends. These distributions are called bell-shaped. The most important bell-shaped distribution is the *normal distribution*. The graph of a normal distribution is sometimes called a *normal curve*, or sometimes a bell curve.

The normal distribution is the most important distribution in mathematics, science, and many other fields. It serves as a very good approximation for many other distributions, including the binomial distribution for large *n*.

Equation of the normal curve:

If a random variable follows a normal distribution (i.e., if the variable is normally distributed) with mean μ and standard deviation σ, then the probability of the variable taking on the value x is given by the following equation:

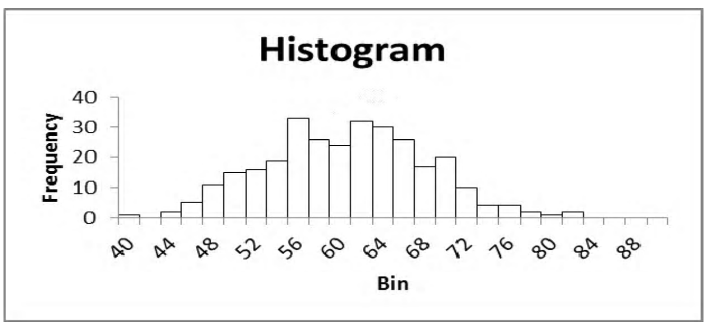
$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

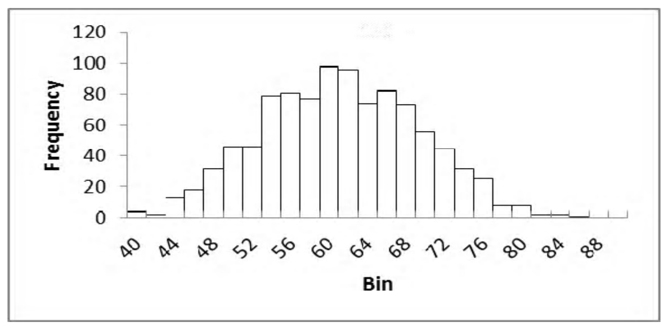
no need to know this!

(You do not need to memorize this equation!)

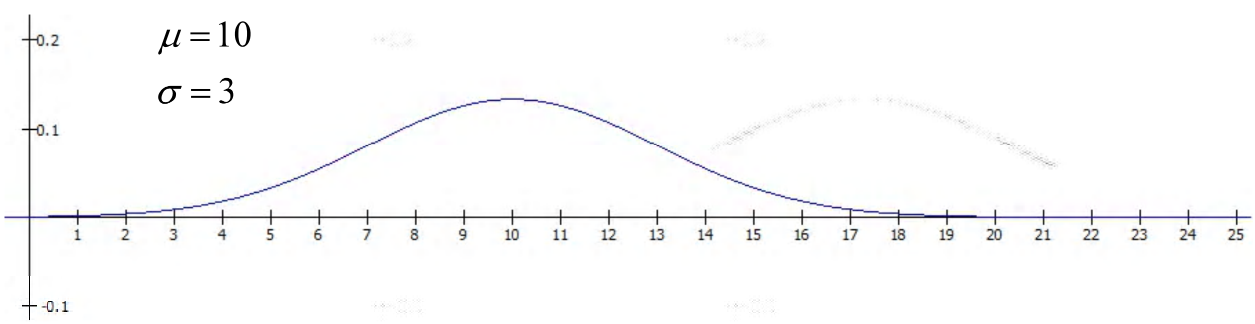
Example 2: Here are the histograms resulting from using Excel to generate random numbers from a normal distribution with mean 60 and standard deviation 8.

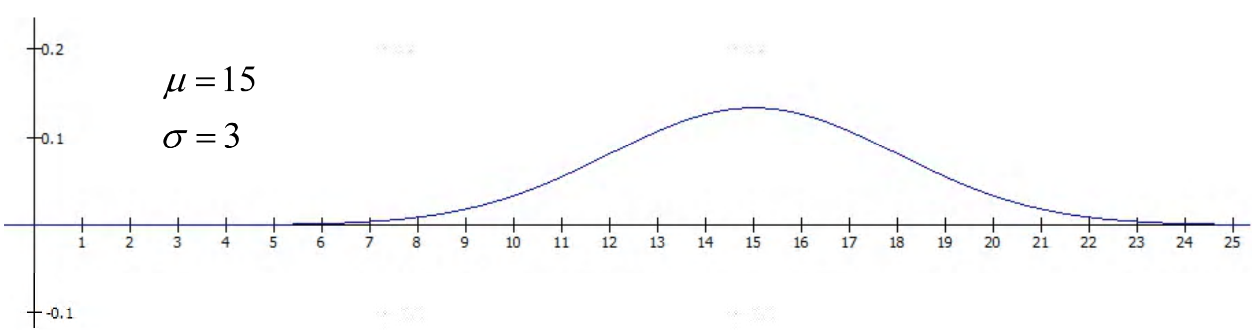
$n = 300$ $n = 1000$

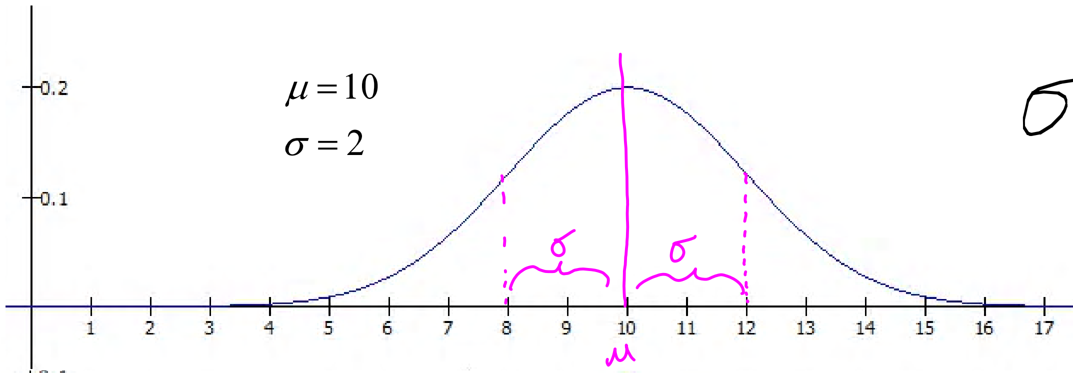




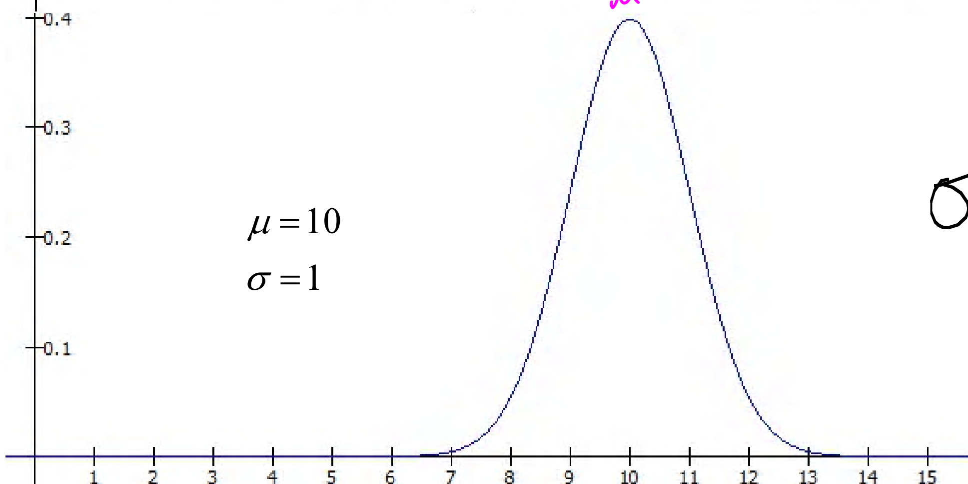
Examples of normal curves:



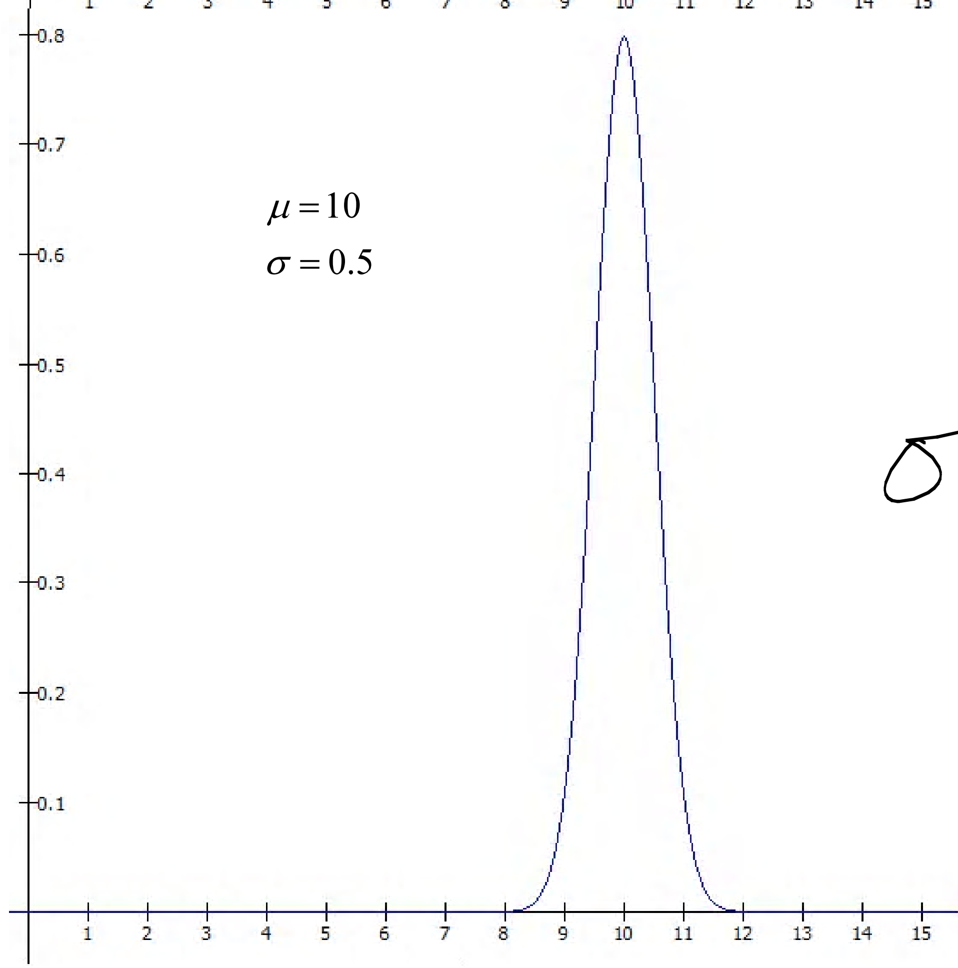




$\sigma = 2$



$\sigma = 1$



$\sigma = 0.5$

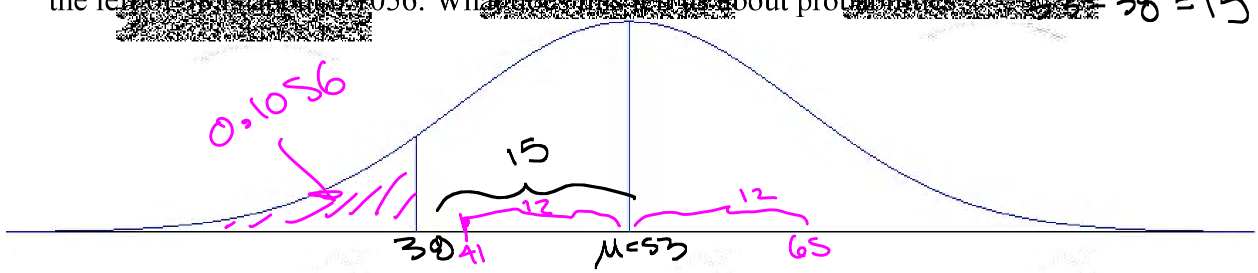
Normal Curve Properties:

1. A normal curve is bell-shaped and symmetric about a vertical line (called its axis of symmetry).
2. The mean is at the axis of symmetry.
3. The shape is completely determined by the mean μ and the standard deviation σ.
   - Small σ: tall, narrow curve
   - Large σ: flat, broad curve
   - Smaller μ: further to the left
   - Larger μ: further to the right
4. The area between the curve and the horizontal axis is always 1. (This corresponds to the fact that all the probabilities in a distribution must add up to 1.)
5. Regardless of the shape,
   - 68.3% of the area is within one standard deviation from the mean.
   - 95.4% of the area is always within two standard deviations from the mean
   - 99.7% of the area is always within three standard deviations from the mean. (Remember the Empirical Rule!)
6. The inflection points (where the graph changes from "concave down" to "concave up" are one standard deviation on either side of the mean (so at $x = \mu + \sigma$ and $x = \mu - \sigma$).

Areas under the normal curve:

One of the many remarkable things about normal curves is the use of area under a normal curve. For any normal curve, with any mean and standard deviation, the areas under the curve areas (and thus the probabilities) can be determined easily by using only one table.

[redacted]e curve to the le[redacted]056. Wh[redacted]bout prob[redacted]38 = 15



a) What percentage of observations are less than 38? $\boxed{10.56\%}$

b) What percentage of observations are greater than 38?

$1 - 0.1056 = 0.8944 \Rightarrow \boxed{89.44\%}$

c) What percentage of observations are greater than 53? $\boxed{50\%}$

d) What percentage of observations lies between 41 and 65?

$\boxed{68.3\%}$

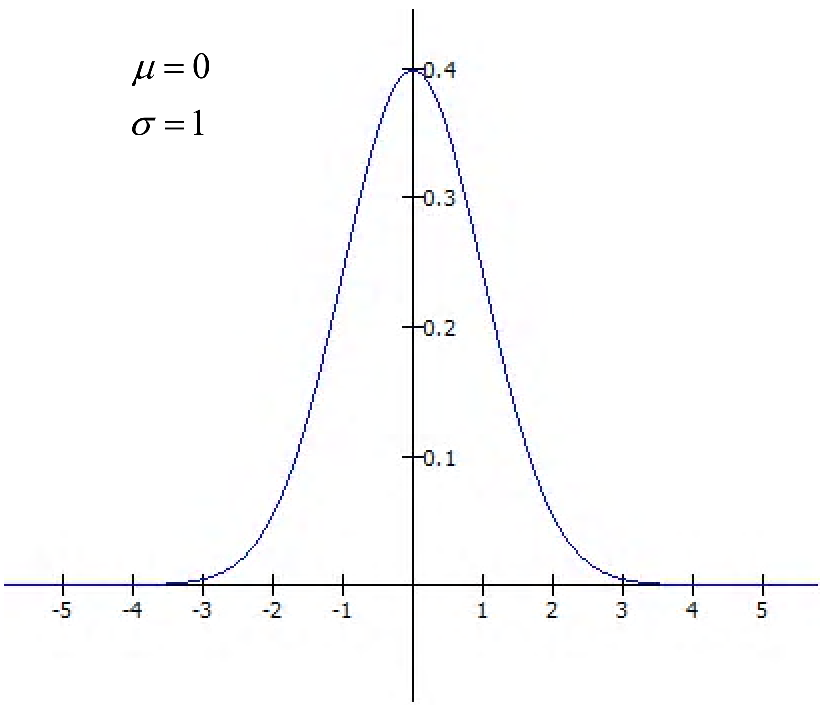
Areas under the normal curve:

One of the many remarkable things about normal curves is the use of area under a normal curve. No matter what the shape of a normal curve, specific areas (and thus probabilities) can be determined easily by using only one table. We *standardize* the values of the variable, by converting them to equivalent values on the *standard normal curve*.

The standard normal curve:

The *standard normal curve* is the normal curve with mean 0 and standard deviation 1. Other normal curves are related to the standard normal curve in this way: the area under any normal curve from μ to $\mu + z\sigma$ corresponds to the area under the standard normal curve from 0 to *z*.

Graph of the standard normal curve:



To standardize the values of a normally distributed variable, we convert them to *z*-scores.

<u>Recall</u>: The *z*-score of a data point is its distance from the mean, measured in standard deviations.

Standardizing the values of a normal distribution:

In a normal distribution with mean μ and standard deviation σ, where x is a data value, the *z*-score is

$$z = \frac{x-\mu}{\sigma}.$$

The area under a normal curve between $x = a$ and $x = b$ is the same as the area under the standard normal curve between the *z*-score for *a* and the *z*-score for *b*.

Example 4: Consider a normal distribution with mean 10 and standard deviation 3.

$\mu = 10, \sigma = 3$

a) What is the *z*-score corresponding to $x = 28$?

28 is 6 SDs above 10, so z-score is +6

or, use the formula: $z = \frac{x-\mu}{\sigma} = \frac{28-10}{3} = \frac{18}{3} = \boxed{6}$

b) What is the *z*-score corresponding to $x = 7$?

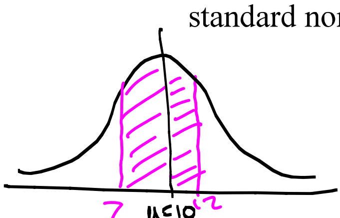
7 is 1 SD below 10, so $z = -1$.

OR $z = \frac{x-\mu}{\sigma} = \frac{7-10}{3} = \frac{-3}{3} = \boxed{-1}$

c) What is the *z*-score corresponding to $x = 20$?

$z = \frac{x-\mu}{\sigma} = \frac{20-10}{3} = \frac{10}{3} = 3\frac{1}{3} = \boxed{3.33}$

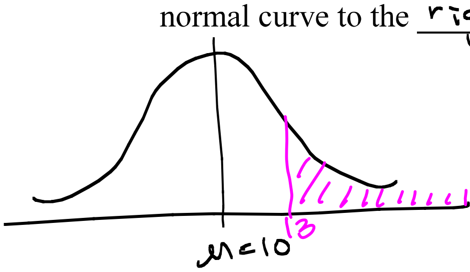
d) The percentage of observations that lie between 7 and 12 is equal to the area under the standard normal curve between ___−1___ and ___0.67___.



we know the z-score for 7 is −1.

z-score for 12: $z = \frac{x-\mu}{\sigma} = \frac{12-10}{3} = \frac{2}{3} \approx 0.67$

e) The percentage of observations that lie above 13 is equal to the area under the standard normal curve to the ___right___ of ___1___.



z-score for 13 is $z = 1$.