Probability Interlude: Normal Distributions (in Investigation 1.7)

We have studied discrete random variables, especially the binomial distribution. Now we will study continuous random variables:
- A random variable is continuous if it can take on any value in an interval of values.
- The probability distribution of a continuous random variable is determined by its probability density function (pdf).
  - This density function must remain non-negative for all values.
  - The total area under the density function must equal 1.
  - The probability that the random variable falls between any two values is given by the area under the density function between those two values.

We will study the most important type of continuous random variable: normal distributions.
- Many continuous random phenomena follow a bell-shaped curve.
- A random variable that follows a symmetric, bell-shaped curve is said to have a normal distribution.
  - The pdf of a normal distribution is: \( f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \).
  - This pdf cannot be integrated in closed form.
  - A normal distribution is characterized by 2 values: mean (\( \mu \)) and standard deviation (\( \sigma \)).
    - The mean \( \mu \) indicates where the center/peak of the distribution is.
    - The SD \( \sigma \) is the distance from the mean to the inflection points (where the curvature changes).

Normal Practice
For each of the three normal curves graphed below, take a guess for its mean and standard deviation:

- **Empirical rule:** With any normal distribution:
  - About 68% of the data falls within one standard deviation of the mean.
  - About 95% of the data falls within two standard deviations of the mean.
  - About 99.7% of the data falls within three standard deviations of the mean.
Example: Comparing SAT and ACT scores
Suppose that the distribution of scores on the SAT exam is symmetric and mound-shaped with mean 1500 and SD 240, while the distribution of scores on the ACT exam is symmetric and mound-shaped with mean 21 and SD 6.

a) About 95% of SAT scores fall between what two values?

b) About 95% of ACT scores fall between what two values?

c) Suppose that Bobby scores 1740 on the SAT. About what percentage of SAT takers scored higher than Bobby? Explain.

d) Suppose that Kathy scores 30 on the ACT. Who has done better compared to their peers – Bobby or Kathy? Explain.

e) Suppose that Peter scores 1020 on the SAT. About what percentage of SAT takers scored higher? Explain.

f) Suppose that Kelly scores 12 on the ACT. Who has done better compared to their peers – Peter or Kelly? Explain.

- A standard score (also called a z-score) for an observation is found by subtracting the mean and then dividing by the standard deviation.
  - Standardization determines number of standard deviations away from the mean.

g) Calculate the standard scores for Bobby, Kathy, Peter, and Kelly. Who has the highest? Who has the lowest?

Bobby: Kathy:

Peter: Kelly:
Normal Probability Calculations

- A normal curve is a probability density function, so the total area under every normal curve is one (or 100%). The area under the curve over a certain interval of values indicates:
  - The proportion of values in that interval
  - The probability that a randomly selected observational unit will be in that interval

- To find the area under the curve over a certain interval, we use technology
  - R: iscamnormprob (see page 63), iscaminvnorm (see page 74)
  - Applet: Normal Probability Calculator
  - Standard normal probability table

Example: Birthweights
Birthweights of babies born in the United States can be modeled by a normal distribution with mean 3250 grams and standard deviation 550 grams.

a) Draw a well-labeled sketch of this distribution.

b) Babies who weigh less than 2500 grams are classified as “low birth weight.” Shade the area corresponding to the probability that a randomly selected baby is low birth weight. Then make an educated guess for the value of this probability.

c) Calculate the \( z \)-score corresponding to 2500 grams. Also interpret what this \( z \)-score reveals.

d) Determine the probability that a randomly selected baby is of low birth weight. Is the probability close to what you guessed earlier?

e) Determine the proportion of babies who weigh more than 10 pounds (4536 grams) at birth.

f) Determine the probability that a randomly selected baby weighs between 3000 and 4000 grams at birth.
g) How little would a baby have to weigh to be among the lightest 2.5% of all newborns? Draw a sketch to support your answer, and also report and interpret the relevant z-score.

h) How much would a baby have to weigh to be among the heaviest 10% of all newborns?

**Example: Candy Bar Weights**
Suppose that the wrapper of a certain candy bar lists its weight as 2.13 ounces. Naturally, the weights of individual bars vary somewhat. Suppose that the weights of these candy bars vary according to a normal distribution with mean \( \mu = 2.20 \) ounces and standard deviation \( \sigma = 0.04 \) ounces.

a) What proportion of candy bars weigh less than the advertised weight?

b) If the manufacturer decides that it’s unacceptable to have 4% of all candy bars weigh less than the advertised weight, they might want to adjust the production process so that only 1 candy bar in 1000 weighs less than advertised. What should the mean of the actual weights be (assuming that the standard deviation of the weights remains 0.04 ounces)? Is this more or less than before? Explain why this makes sense.

c) If the manufacturer does not want to add weight to the candy bars (because this costs money), an alternative is to adjust the SD of the weights in the production process. If the mean weight remains at 2.20 ounces but only 1 candy bar in 1000 weighs less than the advertised weight, how small does the standard deviation of the weights need to be? Is this smaller or larger than before? Explain why this makes sense.

d) If the manufacturer wants to save more money by adjusting the production process so that the mean is reduced to 2.15 ounces but only 1 candy bar in 1000 weighs less than the advertised weight, how small does the standard deviation of the weights need to be?