1. The Binomial Distribution and the Bell Curve

2. You should be familiar with Pascal's Triangle and the Binomial Theorem. You should be familiar with the basics of probability including binomial probabilities and random variables. In this lesson, we will use the bell curve to compute the percentage of a population within a given range and calculate confidence intervals.

3. The bell curve is important in the study of probability and statistics because it appears in a wide variety of applications. We will present its development here as the limit of a large number of coin flips. The bell curve has several important features, such as the symmetry of the left and right tails, and it's shape that has most of the possibilities near the center, with a decreasing likelihood as you move away from the center toward the tails.

4. (a) We begin with a single coin flip and count the number of heads. This distribution is just about as far from the bell curve as you can get, instead of having most of the possibilities in the center, it has the only two possibilities at the extremes, either 100% heads or 100% tails. An important fact about the bell curve is that even this most extreme case, if repeated often, will converge to the bell curve.

   (b) When we flip two coins, and count the number of heads, we begin to see the central tendency. We have a 50% chance of winding up in the middle.

   (c) When we flip three coins, we start to fill up the middle section with more and more bars.

   (d) At four flips, the extremes, no heads and all heads are beginning to disappear.

5. (a) We proceed to five flips,

   (b) and 10 flips. The central bar which represents 5 heads is now getting thin. The probability of getting exactly 50% heads is shrinking, but most of the probability is gathered close to this central average.

6. At 1000 coin flips, the bars are now too thin to see individually, and the curve looks smooth. The extremes have all but disappeared.

7. (a) The limit of this process is called the bell curve because its shape looks like a bell. It is also called the Gaussian distribution, after its discoverer, Karl Gauss. It is also known as the normal distribution, normal in the sense that it appears so frequently in the study of natural phenomena, that its appearance is normally expected. It is defined by two parameters, the first is the average, also called the mean, denoted by the Greek letter μ. Say for instance the we measured the weight of ripe Honeycrisp apples growing from a particular tree. μ would be the average weight.

   (b) We would expect half of the apples to be below average. If X is the weight of a randomly selected apple, the probability that X is below average is $1/2$.

8. (a) Of course by the symmetry of the bell curve, the probability that X is above average is also $1/2$.

(b) The other parameter for the bell curve is the amount that measurements spread out from the mean. This distance is called the standard deviation and is denoted by the Greek letter σ. Otherwise, all bell curves have the same shape. They are completely determined by the mean and the standard deviation.

(c) Bell curves may differ in their mean. The average height of an adult female in the United States is 162 cm, while the average height of a basketball player in the WNBA is 183 cm. The WNBA heights have their mean value shifted to the right.

(d) It is possible for two bell curves to have the same mean, but to be different because one has a smaller standard deviation, making the bell thinner.

(e) Once the mean and standard deviation are known, every bell curve has some standard values. 68% of the population falls within one standard deviation of the mean.

9. (a) If we include the left tail, 84% of the population falls below one standard deviation above the mean.

   (b) This implies that the remaining 15.87% of the population is more than one standard deviation above the mean.

   (c) By symmetry, 15.87% of the population falls below μ minus the standard deviation.

10. (a) These value come from the standard bell curve, also called the standard normal distribution. The standard normal distribution has a mean of 0 and a standard deviation of 1, so that the value being looked up is the number of standard deviations above the mean. For bell curves with other values of μ and σ, we can convert to the standard normal distribution by computing the z-score. The z-score is the number of standard deviations above the mean. To compute the z-score, subtract the mean from the measured value X. This will give you the amount that X is above the mean. Divide that amount by the standard deviation, and z is the number of standard deviations above the mean.

    (b) For example, perhaps we wish to know the percent of American adult females that are shorter than 148 cm. The mean is 162 cm and the standard deviation is 7 cm.

    (c) 148 cm is 14 cm below the mean, which is 2 standard deviations below the mean. The z-score is -2.

11. (a) Many sources have tables for the standard normal distribution. To each z-score is associated the probability that a randomly chosen member of the population is below that z-score. For example, 2.275% of the population is below a z-score of -2. As we would expect, the fraction of the population that is below $z = 0$, which is the fraction of the population below average is 1/2.

    (b) Here is the other half of the table for positive z-scores. Many sources only give this half, ignoring the negative half, which can be found by using the symmetry of the bell curve.

12. (a) Let's return to the height example. What fraction of the population is shorter than 148 cm?

(b) First we compute the z-score.

(c) Then we look up the probability in the table based on the z-score.

13. (a) Here is another example.

    (b) First, we compute the z-score for the value $X = 14$, 14 is 0.6 standard deviations above the mean of 11.

    (c) We then use the table to find the percent of the population below the z-score of 0.6. The table only gives the population below the line.

    (d) However, we weren't asked about the value below 14, we wanted to know the probability of X being above 14, which is the portion of the population that is NOT below 14. To get the answer, we subtract .72575 from 1 to get a probability of 27.425%

14. (a) A related idea is that of a confidence interval. We begin with a set percent of the population that we want to include, and then work backwards to the measured values. Most of the population is in the middle, we wish to chop off the tails by finding upper and lower bounds.

    (b) We typically choose an interval which is centered on the mean. so to include 95% of the population in the interval, we wish to chop off 2.5% from each tail.

    (c) The z-score of -1.96 corresponds to 2.5% of the population below that cutoff, so an interval with z-scores between -1.96 and 1.96 will include 95% of the population.

    (d) We need to convert that z-score back to the actual values. The lower bound should be 1.96 standard deviations below average, which would be at 80.4

15. The upper bound should be 1.96 standard deviations above average, which would be at 119.6

16. Therefore the confidence interval that contains 95% of the population runs from 80.4 to 119.6

17. These are the standard z-scores for the common confidence intervals.

18. Here is another example. We wish to design a bicycle that is adjustable so that 99% of the adult female population can ride comfortably. The z-scores for the 99% confidence interval are ± 2.576 standard deviations from the mean. We will be very close to 99% of the population being able to ride our bike if it adjusts from 144 to 180 cm.

19. (a) Here is a final example. We want to be sure we have enough working parts available, knowing that our supplier isn't very good. On average, a shipment will have 700 working parts, but there is variability. We wish to find an interval that covers 90% of the shipments.

    (b) The z-scores for the 90% confidence interval are ± 1.645 standard deviations, so we find the values that are 1.645 standard deviations of 14.5 from the mean of 700. 90% of the shipments will have between 676 and 724 working parts.