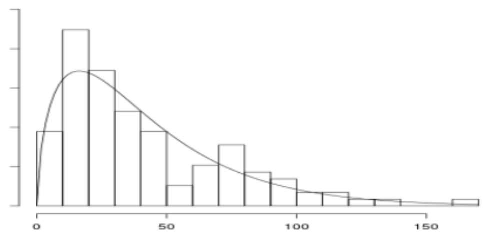
CHAPTER 3 - THE NORMAL DISTRIBUTION

1. Density Curves

We begin this chapter by considering **Density Curves**. We can imagine that a density curve, is an approximation of a histogram. Because of this, density curves tell us what the distribution of a variable looks like.

Figure 1. Histogram and Density Curve

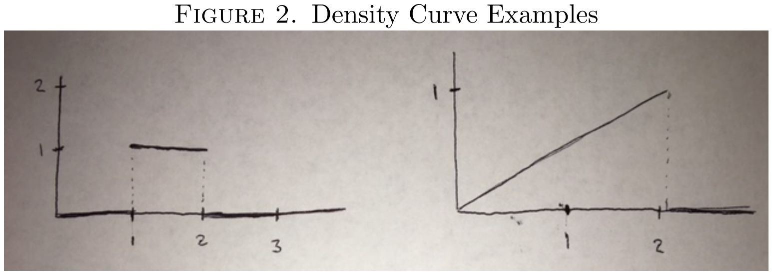


Density curves have some very important properties. These properties will come up again and again, so it is important to know them!

(1) A density curve can never be negative (in the vertical direction).
(2) The area underneath a density curve must always be equal to 1.

For example, check that both of the drawings below are valid density curves. In the previous

Figure 2. Density Curve Examples

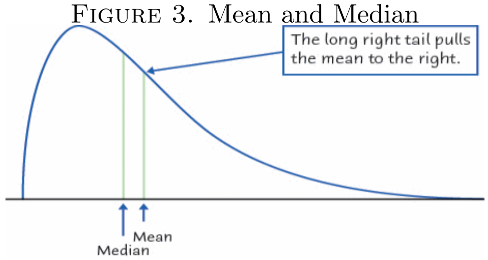


chapter, we talked about mean and median as measures of center. When we are talking about density curves, here is what you should remember.

(1) **Median** - This is the point which splits the area under the curve into equal parts (50% - 50%).
(2) **Mean** - If the density curve was made out of metal, the mean would be the balancing point.

In the example below, we look at the mean and median for a *right skewed* density curve. Notice how the mean is greater than the median in this case because the mean gets pulled further into the tail.

Figure 3. Mean and Median

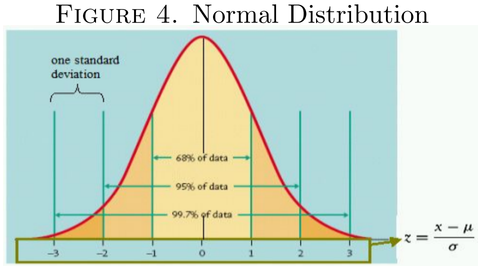


Question: Is the mean or the median larger when the density curve is symmetric? What if it is skewed left?

2. Normal Distributions and the 68-95-99.7 Rule

Now that we have a good understanding of density curves in general, we can look at a very important group of density curves called Normal Distributions. Since the Normal Distribution is

Figure 4. Normal Distribution



a density curve, it must have the two properties discussed earlier. It is hard to check that the area under the curve is equal to 1, but using Calculus we can see that it is true!

The Normal Distribution depends on two parameters. If we know both of them, then we know everything about the shape of the density curve. The parameters are:

(1) The **mean** of the density curve which we denote μ. This is the greek letter 'm' and is pronounced mu. It is the balancing point of the dsitribution. Since Normal density curves are symmetric, it is the exact center and also the Median.
(2) The **standard deviation** of the density curve which we denote σ. This is the greek letter 's' and is pronounced sigma. It is a representation of the *spread* of the density curve. It is the distance from the center of the distribution to the *inflection points* of the curve.
Imagine you are sitting on a sled at the top of the density curve. If you begin to slide down, you will go faster and faster. You reach a point where, while you are still going downhill, the ground **begins** *to level out. That point is called the inflection point.*

In addition to the two usual properties, we notice that Normal Distributions are always single peaked and symmetric, sometimes they are called bell curves. Normal Distributions also follow the $68 - 95 - 99.7$ Rule. This can be seen in Figure 4 above. No matter what the values of μ and σ, the following is true:

68% of the area is contained between $\mu - \sigma$ and $\mu + \sigma$
95% of the area under the curve is contained between $\mu - 2\sigma$ and $\mu + 2\sigma$
99.7% of the area under the curve is contained between $\mu - 3\sigma$ and $\mu + 3\sigma$.

Example 1. Assume that the number of eggs laid per year by a Hawksbill Sea Turtle is Normally Distributed with $\mu = 200$ and $\sigma = 25$. *Note: Sometimes, we write this as N(200, 25)*

- Draw the normal distribution with $\mu = 200$ and $\sigma = 25$.

a. What proportion of turtles lay between 175 and 225 eggs per year?

b. What proportion of turtles lay less than 200 eggs per year?

c. What *percent* of turtles lay more than 225 eggs per year?

d. What *percent* of turtles lay more than 175 eggs but less than 200 eggs per year?

3. Finding Z-Scores

The $68 - 95 - 99.7$ Rule is a nice starting point, and gives us good estimates in some situations. But consider the following question: *What proportion of turtles lay less than 168 eggs per year?* We cannot use the 68-95-99.7 Rule, because 168 doesn't fall on one of the magic numbers. As long as we know μ and σ, we can still find the exact percent but it's a bit tricky. The exact proportions can be calculated and put into a table for us to use, but if we make a new table for every value of μ and σ, we will have to deal with thousands (or more) different tables. The solution is called standardizing the variable, or finding the z-score.

Continuing with the previous example, let's say that the number of eggs laid by a randomly selected turtle is x. The variable x has a mean of 200, so if we say $z = x - 200$, then z should have a mean of 0. But z still has $\sigma = 25$. However if we let $z = \frac{x-200}{25}$, then z should have a mean of 0, and a standard deviation of 1. In general the formula for z looks like this

$$z = \frac{x - \mu}{\sigma} \sim N(0, 1)$$

The second part of the above equation just tells us that z is normally distributed with $\mu = 0$ and $\sigma = 1$. Notice a few things here

If z = 0, then the variable x is exactly average.
if z > 0 then the variable x is above average.
if z < 0 then the variable x is below average.
If z is close to zero, then x is close to average.

Finding a z-score contains the same information as the original x, but now we can look at a table to find the proportion.

Example 2. We find the number of eggs laid by 4 turtles. Fill in the table (calculate the z-score) for each value of x. Then plot both the distributions side by side. You should notice that x and z are in the same location for their respective density curve.

x_i	z_i
200	
175	
270	
168	

Now we can consider problems like we were considering earlier. **Example 3.**

a. What proportion of turtles lay less than 168 eggs per year?

b. What proportion of turtles lay more than 270 eggs per year?

In the above problems, you were given a value of x, and asked to find the proportion. What if you are given a percent and asked to find the corresponding x value? Here are the steps:

1. Find the proportion *inside* Table A, and get the corresponding z-score from the border.
2. Convert this z back to x by using the equation $x = z\sigma + \mu$.

Example 4. If the number of eggs laid by a turtle is in the bottom 1 percent, you may want to examine that turtle to see if she is sick. How many eggs would a sea turtle lay if 99% of turtles lay more than her?

How many eggs would a turtle have to lay, in order to be in the top 34%?

4. Comparing Variables with Z-Scores

There is another way that Z-Scores can come in handy. Since they convert x variables to a common scale, they can be used for comparison.

Example 4. Timothy takes the ACT, and scores a 27. His twin brother Jimothy takes the SAT and scores a 707. Jimothy claims that since 707 is more than 27, he is smarter than his brother. Is he correct?

ACT scores are normally distribtued with $\mu = 20.8$ and $\sigma = 4.8$. SAT scores are normally distributed with $\mu = 606$ and $\sigma = 98$.

5. Review

Essentially, there are two types of problems we can give you.

- You are given a value (x), and asked to find a percent.
- You are given a percent, and asked to find a value (x).

And there are two ways to solve these problems.

- Sometimes, we can use the 68-95-99.7 Rule.
- If the above rule doesn't apply, we must use z-scores and the Standard Normal Table.
  - The strategy is always to get z.
  - If we are given a value of x, use the equation (1) to find z and then the table to find percent.
  - If we are given a percent, use the table to find z and then the equation (2) to find x.

The two equations are given below.

$$z = \frac{x - \mu}{\sigma} \tag{1}$$

$$x = z\sigma + \mu \tag{2}$$