§1 What Is The Normal Distribution

Let’s look at an example first. Suppose a researcher selects a random sample of 100 adult men, measures their heights, and constructs a histogram for the data. It may look like the histogram below.

What can we tell from this histogram? Most of the heights are centered around the middle, and decrease as you go towards the edges. This makes sense. Most adult men are of average height (say, 5 feet 10 inches or so) and then there are men that are shorter, and men that are taller. When we say average, though, we expect that most of the men are around that height, while the number of men who are shorter than the average and the number of men who are taller than the average decreases in number. How many men do you know who are over 7 feet tall? Or under 5 feet tall?

What we notice, though, is that as the sample size increases, the shape of histogram will not change that much at all. Take a look at the following diagram.
If we could measure the heights of all adult males, then the histogram at the bottom right may represent these heights. Any histogram that looks like this follows what we call the normal distribution. It is also called the bell-shaped curve of the Gaussian distribution (after famous German mathematician Carl Gauss). Note that in the normal distribution, the mean, median and mode are all the same and are located at the center of the distribution.

There are many phenomena that follows the normal distribution. Heights and weights of adult males, IQ’s, prices paid for new cars, etc… In these distributions, the data items tend to cluster around the mean. The more an item differs from the mean, the less likely it is to occur.

§2 The Standard Deviation And Percentages

The standard deviation is important because it explains how much of the data falls around the mean. There is a simple rule to remember the numbers: 68-95-99.7. This means that 65% of the data values fall within one standard deviation of the mean (above and below). 95% of the data values fall within two standard deviations of the mean, and 99.7% of the data values fall within three standards deviations from the mean. These numbers are not exact but are very close. We will find the exact values later.

**THE 68–95–99.7 RULE FOR THE NORMAL DISTRIBUTION**

1. Approximately 68% of the data items fall within 1 standard deviation of the mean (in both directions).
2. Approximately 95% of the data items fall within 2 standard deviations of the mean.
3. Approximately 99.7% of the data items fall within 3 standard deviations of the mean.
Note that as we move away from the mean, the curve falls rapidly, and then gradually to the horizontal axis. The tails never quite touch the horizontal axis. The range of the normal distribution is infinite. No matter how far out from the mean we move, there is always a chance, or probability, of a data item occurring even further out.

Let's look at an example. Say the heights of adult males is normally distributed with a mean of 70 inches and a standard deviation of 4 inches.

Then we can see that a height of 74 inches is one standard deviation above the mean and a height of 78 inches is two standard deviations above the mean. Similarly, a height of 66 inches is one standard deviation below the mean and a height of 62 inches is two standard deviations below the mean. The distribution can be illustrated by the figure below.

When working on any problem dealing with the normal distribution, the above graph is the first thing you should always do.

Note the relationship between the standard deviation and percentages. 68% of adult males have a height between 66 inches and 74 inches. 95% of adult males have a height between 62 inches and 78 inches. Because this graph is symmetric, we can also say that 34% of adult males have a height between 70 inches and 74 inches. We can also say that 2.5% of adult males have a height below 62 inches.

PRACTICE

1) The mean score on a standardized math test is 640 points with a standard deviation of 30 points. Draw the normal curve representing this distribution with scores for one, two, and three standard deviations above and below the mean.
§3 The Standard Deviation And Z-Scores

In the normal distribution, a z-score simply describes how many standard deviations a data item lies above or below the mean. The z-score can be obtained by using the following:

\[
\text{z-score} = \frac{\text{data item} - \text{mean}}{\text{standard deviation}}
\]

Data items above the mean have a positive z-score. Data items below the mean have a negative z-score.

What is the z-score for the mean? Based on the formula above, the mean has a z-score of zero!

PRACTICE

2) The weight of newborn infants is normally distributed with a mean of 7 pounds and a standard deviation of 0.8 pounds. Make the bell curve showing the weights that are one, two, and three standard deviations above and below the mean. What is the z-score for an infant who weighs 6 pounds? What is the z-score for an infant who weighs 9 pounds?

Once you are given the mean and standard deviation of a normal distribution, you should be able to find the z-score for any value. You should also be able to go backwards.

Say the IQ’s are normally distributed with a mean of 100 and standard deviation of 16. What IQ corresponds to a z-score of -1.5? What IQ corresponds to a z-score of 2.05?

To answer these questions, simply use the formula! Note that we ARE NOT finding the z-score here. We want to find the data item given the z-score. So we plug in the given values and work backwards.

The equation for the first part becomes \(-1.5 = \frac{x-100}{16}\). Solve for x to get that \(x = 76\).

The equation for the second part becomes \(2.05 = \frac{x-100}{16}\). Solve for x to get that \(x = 132.8\)

PRACTICE

3) The SAT has a mean of 500 and standard deviation of 100. The ACT has a mean of 18 and a standard deviation of 6. Both tests are normally distributed. Suppose you scored a 550 on the SAT and a 24 on the ACT. On which test did you have the better score? Explain why.