Print

Understanding the Normal Distribution

For continuous data, a histogram is the best way to see the distribution of sample data.

Here, we see histograms for variables with four different distributions: bell-shaped, right-skewed, left-skewed, and uniform.

When we graph sample data using a histogram, we are usually trying to understand or estimate the shape of the distribution of the underlying population. That is, we want to understand the underlying continuous probability distribution.

This bell-shaped histogram is centered at zero, and most of the observations fall within plus or minus 3 units from the center of the distribution.

This distribution can also be described as mounded in shape and symmetric about the center.

This is an example of sample data that are approximately normally distributed. From these sample data, you can estimate the distribution of the population.

Here, we have fit a normal curve to the data. The normal curve fits the data well. This is an indication that the underlying population distribution might be approximately normal.

The normal distribution is one of the most common reference distributions in statistics, with many useful mathematical properties. One property of the normal distribution is that the shape depends entirely on two independent values: the mean (μ) and the standard deviation (σ).

As you learned in the previous video, the population mean (μ) is the center of the distribution and the standard deviation (σ) measures the spread or dispersion of the distribution.

The larger the standard deviation, the wider the distribution.

These data were actually simulated using a normal distribution with a population mean of 0 and a population standard deviation of 1.

This particular normal distribution has a name: the standard normal distribution.

These are all normal distributions, with different means and different standard deviations.

The total area under the normal curve is 1. As a result, you can calculate probabilities for values of a normally distributed variable by calculating the area under the curve.

Approximately 68% (or 68.27%) of the area under the normal curve falls within plus or minus 1 standard deviation of the mean.

This means that approximately 68% of the data values for a variable that is normally distributed will fall within plus or minus 1 standard deviation of the mean.

Approximately 95% (or 95.4%) of the area under the normal curve falls within plus or minus 2 standard deviations of the mean.

And 99.73% of the area under the normal curve falls within plus or minus 3 standard deviations of the mean.

Why is this important?

If your data are normally distributed, fewer than 5% of the observations will fall more than 2 standard deviations from the mean, and approximately 0.27% of the observations will fall more than 3 standard deviations from the mean. It is very unlikely that you will observe a value that falls more than 3 standard deviations from the mean, unless something unusual has happened.

Because the normal distribution has many useful mathematical properties, statistical procedures often assume the normal distribution.

In the next video, you learn how to see if your data are normally distributed.

Statistical Thinking for Industrial Problem Solving

Close