

## Statistics Translated: A Step-by-Step Guide to Analyzing and Interpreting Data

### What Is Statistics?

Statistics is a branch of applied mathematics that involves the collection, description, analysis, and inference of conclusions from quantitative data. The mathematical theories behind statistics rely heavily on differential and integral calculus, linear algebra, and probability theory.

Statisticians, people who do statistics, are particularly concerned with determining how to draw reliable conclusions about large groups and general events from the behavior and other observable characteristics of small samples. These small samples represent a portion of the large group or a limited number of instances of a general phenomenon.

**Key Takeaways** Statistics is the study and manipulation of data, including ways to gather, review, analyze, and draw conclusions from data.

The two major areas of statistics are descriptive and inferential statistics.

Statistics can be communicated at different levels ranging from non-numerical descriptor (nominal-level) to numerical in reference to a zero-point (ratio-level).

A number of sampling techniques can be used to compile statistical data including simple random, systematic, stratified, or cluster sampling.

Statistics are present in almost every department of every company and are an integral part of investing as well.

### 2:00 Statistics

#### Understanding Statistics

Statistics are used in virtually all scientific disciplines such as the physical and social sciences, as well as in business, the humanities, government, and manufacturing. Statistics is fundamentally a branch of applied mathematics that developed from the application of mathematical tools including calculus and linear algebra to probability theory.

In practice, statistics is the idea we can learn about the properties of large sets of objects or events (a population) by studying the characteristics of a smaller number of similar objects or events (a sample). Because in many cases gathering comprehensive data about an entire population is too costly, difficult, or flat out impossible, statistics start with a sample that can conveniently or affordably be observed.

Two types of statistical methods are used in analyzing data: descriptive statistics and inferential statistics. Statisticians measure and gather data about the individuals or elements of a sample, then analyze this data to generate descriptive statistics. They can then use these observed

## P

characteristics of the sample data, which are properly called "statistics," to make inferences or educated guesses about the unmeasured (or unmeasured) characteristics of the broader population, known as the parameters.

Statistics informally dates back centuries. An early record of correspondence between French mathematicians Pierre de Fermat and Blaise Pascal in 1654 is often cited as an early example of statistical probability analysis.

### Descriptive and Inferential Statistics

The two major areas of statistics are known as descriptive statistics, which describes the properties of sample and population data, and inferential statistics, which uses those properties to test hypotheses and draw conclusions. Descriptive statistics include mean (average), variance, skewness, and kurtosis. Inferential statistics include linear regression analysis, analysis of variance (ANOVA), logit/Probit models, and null hypothesis testing.

### Descriptive Statistics

Descriptive statistics mostly focus on the central tendency, variability, and distribution of sample data. Central tendency means the estimate of the characteristics, a typical element of a sample or population, and includes descriptive statistics such as mean, median, and mode. Variability refers to a set of statistics that show how much difference there is among the elements of a sample or population along the characteristics measured, and includes metrics such as range, variance, and standard deviation.

The distribution refers to the overall "shape" of the data, which can be depicted on a chart such as a histogram or dot plot, and includes properties such as the probability distribution function, skewness, and kurtosis. Descriptive statistics can also describe differences between observed characteristics of the elements of a data set. Descriptive statistics help us understand the collective properties of the elements of a data sample and form the basis for testing hypotheses and making predictions using inferential statistics.

### Inferential Statistics

Inferential statistics are tools that statisticians use to draw conclusions about the characteristics of a population,

drawn from the characteristics of a sample, and to decide how certain they can be of the reliability of those conclusions. Based on the sample size and distribution statisticians can calculate the probability that statistics, which measure the central tendency, variability, distribution, and relationships between characteristics within a data sample, provide an accurate picture of the corresponding parameters of the whole population from which the sample is drawn.

Inferential statistics are used to make generalizations about large groups, such as estimating average demand for a product by surveying a sample of consumers' buying habits or to attempt to predict future events, such as projecting the future return of a security or asset class based on returns in a sample period.

Regression analysis is a widely used technique of statistical inference used to determine the strength and nature of the relationship (i.e., the correlation) between a dependent variable and one or more explanatory (independent) variables. The output of a regression model is often analyzed for statistical significance, which refers to the claim that a result from findings generated by testing or experimentation is not likely to have occurred randomly or by chance but is likely to be attributable to a specific cause elucidated by the data. Having statistical significance is important for academic disciplines or practitioners that rely heavily on analyzing data and research.

#### Understanding Statistical Data

The root of statistics is driven by variables. A variable is a data set that can be counted that marks a characteristic or attribute of an item. For example, a car can have variables such as make, model, year, mileage, color, or condition. By combining the variables across a set of data (i.e. the colors of all cars in a given parking lot), statistics allows us to better understand trends and outcomes.

There are two main types of variables. First, qualitative variables are specific attributes that are often non-numeric. Many of the examples given in the car example are qualitative. Other examples of qualitative variables in statistics are gender, eye color, or city of birth. Qualitative data is most often used to determine what percentage of an outcome occurs for any given qualitative variable, and qualitative analysis often does not rely on numbers. For example, trying to determine what percentage of women own a business analyzes qualitative data.

The second type of variable in statistics is quantitative variables. Quantitative variables are studied numerically

and only have weight when about a non-numerical descriptor. Similar to quantitative analysis, this information is rooted in numbers. In the car example above, the mileage driven is a quantitative variable. However, the number 60,000 holds no value unless it is understood that is the total number of miles driven.

Quantitative variables can be further broken into two categories. First, discrete variables have limitations in statistics and infer that there are gaps between potential discrete variable values. The number of points scored in a football game is a discrete variable because (1) there can be no decimals and (2) it is impossible for a team to score only 1 point.

Second, statistics also makes use of continuous quantitative variables. These values run along a scale - whereas discrete values have limitations, continuous variables are often measured into decimals. When measuring the height of the football players, any value (within possible limits) can be obtained, and the heights can be measured down to 1/16ths of an inch if not further.

Statisticians can hold different titles and positions within a company. According to Glassdoor, the average total compensation for a statistician as of December 2021 was \$98,034. An equally analytical role of data scientist yielded annual compensation of almost \$119,000.

#### Statistical Levels of Measurement

After analyzing variables and outcomes as part of statistics, there are several resulting levels of measurement. Statistics can quantify outcomes in these different ways:

Nominal Level Measurement. There is no numerical or quantitative value, and qualities are not ranked. Instead, nominal level measurements are simply labels or categories assigned to other variables. It's easiest to think of nominal level measurements as non-numerical facts about a variable. Example: The name of the President elected in 2020 was Joseph Robinette Biden, Jr. Ordinal Level Measurement: Outcomes can be arranged in an order, however, all data values have the same value or weight. Although numerical, ordinal level measurements in statistics can't be subtracted against each other as only the position of the data point matters. Often incorporated into nonparametric statistics, ordinal levels are often compared against the total variable group. Example: American Fred Kerley was the 2nd fastest man at the 2020 Tokyo Olympics based on 100-meter sprint times. Interval Level Measurement: Outcomes can be arranged in

order; however differences between data values may now have meaning. Two different data points are often used to compare the passing of time or changing conditions within a data set. There is often no "starting point" for the range of data values, and calendar dates or temperatures may not have a meaningful intrinsic zero value. Example: Inflation hit 8.6% in May 2022. The last time inflation was this high was December 1981. Ratio Level Measurement: Outcomes can be arranged in order, and differences between data values now have meaning. However, there is now a starting point or "zero value" that can be used to further provide value to a statistical value. The ratio between data values now has meaning, including its distance away from zero. Example: The lowest meteorological temperature recorded was -128.6 degrees Fahrenheit in Antarctica.

#### Statistics Sampling Techniques

To gather statistical information, it would often not be possible to gather data from every data point within a population. Instead, statistics relies on different sampling techniques to create a representative subset of the population that is easier to analyze. In statistics, there are several primary types of sampling.

Simple random sampling calls for every member within the population to have an equal chance of being selected for analysis. The entire population is used as the basis for sampling, and any random generator based on chance can select the sample items. For example, 100 individuals are lined up and 10 are chosen at random.

calls for every member within the population to have an equal chance of being selected for analysis. The entire population is used as the basis for sampling, and any random generator based on chance can select the sample items. For example, 100 individuals are lined up and 10 are chosen at random. Systematic sampling calls for a random sample as well. However, its technique is slightly modified to make it easier to conduct. A single random number is generated, and individuals are then selected at a specified regular interval until the sample size is complete. For example, 100 individuals are lined up and numbered. The 7th individual is selected for the sample followed by every subsequent 9th individual until 10 sample items have been selected.

calls for a random sample as well. However, its technique is slightly modified to make it easier to conduct. A single random number is generated, and individuals are then selected at a specified regular interval until the sample size is complete. For example, 100 individuals are lined up and numbered. The 7th individual is selected for the sample followed by every subsequent 9th individual until 10 sample items have been selected. Stratified sampling calls for

more control over your sample. The population is divided into subgroups based on similar characteristics. Then, you calculate how many people from each subgroup would represent the entire population. For example, 100 individuals are grouped by gender and race. Then, a sample from each subgroup will be taken in the proportion of how representative that subgroup is of the population.

calls for more control over your sample. The population is divided into subgroups based on similar characteristics. Then, you calculate how many people from each subgroup would represent the entire population. For example, 100 individuals are grouped by gender and race. Then, a sample from each subgroup will be taken in the proportion of how representative that subgroup is of the population. Cluster sampling calls for subgroups as well. However, each subgroup should be representative of the population. Instead of randomly selecting individuals within a subgroup, the entire subgroup is randomly selected.

Not sure which Major League Baseball player should have won Most Valuable Player last year? Statistics, often used to determine value, is often cited when the award for best player is awarded. Statistics can include batting average, number of home runs hit, and stolen bases.

#### Examples of Statistics

Statistics is prominent in finance, investing, business, and the world. Much of the information you see and the data you are given is derived from statistics, which are used in all facets of a business.

In investing, statistics include average trading volume, 52-week low, 52-week high, beta, and correlation between asset classes or securities.

In economics, statistics include GDP, unemployment, consumer pricing, and inflation, and other economic growth metrics

In marketing, statistics include conversion rates, click-through rates, search quantities, and social media metrics.

In accounting, statistics include liquidity, solvency, and profitability metrics across time.

In information technology, statistics include bandwidth, network capabilities, and hardware logistics.

In human resources, statistics include employee turnover, employee satisfaction, and average compensation relative to the market.

Why Is Statistics Important? Statistics provide the information to educate how things work. Statistics are used to conduct research, evaluate outcomes, develop critical thinking, and make informed decisions. Statistics can be used to inquire almost any field of study to investigate why things happen, when they occur, and whether its reoccurrence is predictable.

What Is the Difference Between Descriptive and Inferential Statistics? Descriptive statistics are used to describe or summarize the characteristics of a sample or data set, such as a variable's mean, standard deviation, or frequency. Inferential statistics, in contrast, employs any number of techniques to relate variables in a data set to one another, for example using correlation or regression analysis. These can then be used to estimate forecasts or infer causality.

Who Uses Statistics? Statistics are used widely across an array of applications and professions. Any time data are collected and analyzed, statistics are being done. This can range from government agencies to academic research to analyzing investments.

## Reference

[IBM SPSS Statistics 26 Step by Step: A Simple Guide and Reference](#)

[Principles of Medical Professionalism](#)