**Probability and Statistics**

**Probability and Statistics** are helpful guides when it comes to studying numbers. Probability helps us figure out how likely things are to happen, like guessing if it will rain. On the other hand, Statistics is about collecting and understanding data, like looking at numbers to learn useful things. Together, they help us make smart decisions and see patterns in the information around us.

This article covers various concepts of probability and statistics. In probability, we will learn definitions, formulas, types of events, Rules of probability, and more topics related to probability.

*Probability and Statistics*

**Table of Content**

* [What is Probability And Statistics?](https://www.geeksforgeeks.org/probability-and-statistics/#what-is-probability-and-statistics)
* [Probability Definition](https://www.geeksforgeeks.org/probability-and-statistics/#probability-definition)
* [Statistics Definition](https://www.geeksforgeeks.org/probability-and-statistics/#statistics-definition)
* [Terms Related to Probability and Statistics](https://www.geeksforgeeks.org/probability-and-statistics/#terms-related-to-probability-and-statistics)
* [Probability and Statistics Formulas](https://www.geeksforgeeks.org/probability-and-statistics/#probability-and-statistics-formulas)
* [Probability Formulas](https://www.geeksforgeeks.org/probability-and-statistics/#probability-formulas)
* [Statistics Formulas](https://www.geeksforgeeks.org/probability-and-statistics/#statistics-formulas)
* [Topics under Probability and Statistics](https://www.geeksforgeeks.org/probability-and-statistics/#topics-under-probability-and-statistics)
  + [Rules of Probability](https://www.geeksforgeeks.org/probability-and-statistics/#rules-of-probability)
* [Types of Event in Probability](https://www.geeksforgeeks.org/probability-and-statistics/#types-of-event-in-probability)
* [Statistics Topics](https://www.geeksforgeeks.org/probability-and-statistics/#statistics-topics)
* [Probability and Statistics for Engineering Mathematics](https://www.geeksforgeeks.org/probability-and-statistics/#probability-and-statistics-for-engineering-mathematics)
  + [Probability in Engineering](https://www.geeksforgeeks.org/probability-and-statistics/#probability-in-engineering)
  + [Statistics in Engineering](https://www.geeksforgeeks.org/probability-and-statistics/#statistics-in-engineering)
* [Probability and Statistics – Solved Examples](https://www.geeksforgeeks.org/probability-and-statistics/#probability-and-statistics-solved-examples)
* [Practice Problems on Probability and Statistics](https://www.geeksforgeeks.org/probability-and-statistics/#practice-problems-on-probability-and-statistics)

**What is Probability And Statistics?**

[**Probability**](https://www.geeksforgeeks.org/probability-in-maths) deals with the likelihood of events occurring, assigning a measure between 0 and 1 to quantify uncertainty. It helps us understand the chances of different outcomes in uncertain situations, such as predicting weather or game results. On the other hand, [**Statistics**](https://www.geeksforgeeks.org/statistics) involves collecting, analyzing, and interpreting data to draw meaningful conclusions. It provides tools to summarize information, identify patterns, and make informed decisions in diverse fields, contributing to our understanding of uncertainty and variability in real-world scenarios.

**Probability Definition**

Probability is a measure of the likelihood or chance of an event occurring. It is expressed as a number between 0 and 1, where 0 indicates an impossible event, and 1 signifies a certain event. Probability is calculated by dividing the number of favorable outcomes by the total number of possible outcomes. In simple terms, it quantifies the likelihood of an outcome in a given set of circumstances, providing a basis for making informed predictions and decisions in various fields, including [mathematics](https://www.geeksforgeeks.org/maths/), [statistics](https://www.geeksforgeeks.org/statistics/), and everyday life.

**Read in Detail:**[Probability Theory](https://www.geeksforgeeks.org/probability-theory)

**Statistics Definition**

Statistics is the branch of mathematics that involves the collection, analysis, interpretation, presentation, and organization of data. It provides methods for making inferences about populations based on samples. In a broader sense, statistics helps to quantify uncertainty and variation in data, enabling researchers, analysts, and decision-makers to draw meaningful conclusions and make informed decisions. It encompasses various techniques, including descriptive statistics to summarize data and inferential statistics to make predictions or test hypotheses about larger populations.

**Read in Detail:**[Statistics in Maths](https://www.geeksforgeeks.org/statistics)

**Terms Related to Probability and Statistics**

* [**Random Experiment**](https://www.geeksforgeeks.org/random-experiment-probability)**:**An experiment is a set of steps that gives clear results. A random experiment is one where you can’t predict the exact result.
* **Outcome:**Outcome means any possible result in a group of results, called a sample space, noted as S. For example, when you flip a fair coin, the sample space is {heads, tails}.
* [**Sample Space**](https://www.geeksforgeeks.org/sample-space-probability)**:**Sample space is the collection of all possible outcomes in an experiment. Like in a coin flip, the sample space is {heads, tails}.
* [**Event**](https://www.geeksforgeeks.org/events-in-probability)**:**An event is any part of a sample space. If an event A happens, it means one of the outcomes in A has occurred. For instance, if event A is rolling an even number on a fair six-sided die, getting 2, 4, or 6 means event A occurred. If you get 1, 3, or 5, event A did not happen.
* **Trial:**A trial is each time you do an experiment, like flipping a coin. In the coin-flipping experiment, each flip of the coin is a trial.
* **Mean:**A random variable’s mean is the average of the values it could have during a random experiment.
* [**Expected Value**](https://www.geeksforgeeks.org/expected-value)**:**The expected value is the mean of a [random variable](https://www.geeksforgeeks.org/random-variable/). For instance, if we roll a six-sided die, the expected value is the [average](https://www.geeksforgeeks.org/average/) of all possible outcomes, which is 3.5.

**Probability and Statistics Formulas**

Some of the common formulas of [Probability](https://www.geeksforgeeks.org/probability-in-maths/) and Statistics are discussed below:

**Probability Formulas**

Probability is the likelihood of an event occurring and is calculated using the following formula:

***P(A) = Number of Favourable Outcomes / Total Number of Possible Outcomes***

*Where:*

*P(A) is the probability of event A.*

*Number of Favorable Outcomes is the count of outcomes where event A occurs.*

*Total Number of Possible Outcomes is the count of all possible outcomes.*

In simple terms, probability is the ratio of successful outcomes to all possible outcomes. The result is a number between 0 (impossible event) and 1 (certain event). It can also be expressed as a percentage by multiplying the result by 100.

**For example,**if you want to find the probability of rolling a 4 on a six-sided die, there is 1 favorable outcome (rolling a 4) out of 6 possible outcomes (1, 2, 3, 4, 5, 6). Therefore,

P(rolling a 4)= 1/6

This formula provides a basic way to express the likelihood of events in a mathematical manner.

**Statistics Formulas**

Some of the common formulas for statistics are discussed below:

**Mean**

The [mean](https://www.geeksforgeeks.org/what-is-mean) is the average of a set of numbers. To find the mean, add up all the numbers in a dataset and then divide by the total number of values.

Mean = Sum of all values / Total number of values

*xˉ=∑xiNxˉ=N∑xi​​*

**W**here,

* xˉ*x*ˉ is the mean,
* ∑xi is the sum of all terms in the data set,
* N is the total number of terms.

**Median**

The median is the middle value in a dataset when it’s arranged in ascending or descending order. If there’s an even number of values, the median is the average of the two middle numbers.

**Median (Odd n)**

*Median = Value at (n+12)(2n+1​)th position*

**Median (Even n)**

*Median=12(Value atn2th position+Value at(n2+1)th position)Median=21​(Value at2n​th position+Value at(2n​+1)th position)*

Where,

* n is the number of values in the data set

**Mode**

The mode is the value that appears most frequently in a [dataset](https://www.geeksforgeeks.org/what-is-dataset/). A dataset may have one mode (unimodal), more than one mode (multimodal), or no mode at all.

**Variance**

Variance measures how spread out the values in a dataset are. It’s calculated by finding the average of the squared differences between each value and the mean.

*Variance= ∑(Each value−Mean) 2 / Total number of values*

*OR*

*σ2=∑(xi–xˉ)2Nσ2=N∑(xi​–xˉ)2​*

Where,

* σ2 is the variance
* ∑(xi− xˉ*x*ˉ)2is the sum of squared differences between each term and the [mean](https://www.geeksforgeeks.org/mean/)
* N is the total number of terms.

**Standard Deviation**

[Standard deviation](https://www.geeksforgeeks.org/standard-deviation-formula) is the square root of the variance. It provides a more interpretable measure of how spread out the values are in comparison to the mean.

*Standard Deviation = √Variance*

*OR*

*σ2=∑(xi–xˉ)2Nσ2​=N∑(xi​–xˉ)2​​*

Where,

* xi represents each term in the data set
* σ2 is the variance,
* √σ2 is the [standard deviation](https://www.geeksforgeeks.org/standard-deviation-formula/).
* xˉ*x*ˉ is the mean

**Topics under Probability and Statistics**

Some important topics under both Probability and Statistics are discussed below:

**Rules of Probability**

The three Major rules of probability we study in probability are:

**I. Addition Rule**

The [addition rule of probability](https://www.geeksforgeeks.org/addition-rule-for-probability) is used when you want to find the probability of at least one of two mutually exclusive events happening. For two mutually exclusive events A and B, the probability of either event occurring (denoted as P(A or B)) is found by adding the individual probabilities of A and B.

If A and B are not mutually exclusive events, the probability of A or B occurring is:

P(A ∪ B) = P(A) + P(B) – P(A ∩ B)

where P(A ∩ B) is the probability of A and B occurring.

If A and B are mutually exclusive events, then

P(A ∪ B) = P(A) + P(B),

since P(A ∩ B) = 0.

**II. Multiplication Rule**

The [multiplication rule of probability](https://www.geeksforgeeks.org/multiplication-theorem) is used when you want to find the probability of two independent events happening together. To find the chance of two events A and B happening at the same time, the multiplication rule is used. If A and B depend on each other, the probability of both events occurring is the product of the probability of A and the conditional probability of B given that A has happened.

P(A ∩ B)=P(A)×P(B∣A)

Here,

P(B∣A) is the likelihood of event B happening when event A has already occurred.

**III. Bayes’ Rule**

[Bayes’ Rule](https://www.geeksforgeeks.org/bayes-theorem) is a formula used to update probabilities based on new evidence. It calculates the probability of an event A happening given the occurrence of another event B. The formula is as follows:

P(A∣B) = P(B)

P(B∣A) × P(A)

Here:

* P(A∣B) is the probability of event A occurring given that event B has occurred.
* P(B∣A) is the probability of event B occurring given that event A has occurred.
* P(A) and P(B) are the probabilities of events A and B occurring, respectively.

**IV. Some Other Rules**

* **Probability is between 0 and 1:**The likelihood of an event ranges from 0 (impossible) to 1 (certain). A probability of 0.5 means an equal chance.
* **The sum of all probabilities is 1:** When you consider all possible outcomes of an event, the total probability is 1. If one outcome has a probability of 0.3, the other outcome (or outcomes) must add up to 0.7 to make 1.
* **Complement Rule:** The probability of an event happening (P(A)) plus the probability of it not happening (P(not A)) equals 1. P(not A) is often written as 1−P(A).

**Types of Event in Probability**

The various [types of events in probability](https://www.geeksforgeeks.org/events-in-probability) are:

* **Simple Event**

A simple event is when an outcome has just one possibility. For instance, in coin flipping, getting heads is a simple event, and getting tails is another. The probability of a simple event is determined by the formula:

*P(Simple Event) = 1 / Total Possible Outcomes*

* **Compound Event**

A compound event involves two or more simple events. For example, flipping a coin twice and getting heads both times is a compound event. The probability of a compound event is found by multiplying the probabilities of its independent simple events.

*P(Compound Event) = P(Event 1) × P(Event 2)*

* **Independent Event**

Independent events are those where the outcome of one event doesn’t affect the outcome of another. Flipping a fair coin is an example; each flip has an equal chance of heads or tails.

* **Dependent Event**

Dependent events are influenced by the outcome of another event. For instance, drawing marbles from a bag without replacement changes the probability for subsequent draws.

* **Complementary Event**

The complement of an event (denoted as A’) includes all outcomes not in event A. If the probability of rolling an even number on a fair six-sided die is event A, then the probability of not rolling an even number (rolling an odd number) is the complement, and it’s calculated as:

*P(Not A) = 1−P(A)*

* **Probability Distribution**

A [probability distribution](https://www.geeksforgeeks.org/probability-distribution) describes how the probabilities of different outcomes are spread across the possible values of a random variable. It provides a comprehensive view of the likelihood of each possible outcome, helping to understand the uncertainty associated with random events. There are two main types of probability distributions:

* [Discrete Probability Distribution](https://www.geeksforgeeks.org/discrete-probability-distribution/#:~:text=A%20discrete%20probability%20distribution%20that%20gives%20the%20number%20of%20events,e%2D%C6%9B%5D%20%2F%20x!)
* [Continuous Probability Distribution](https://www.geeksforgeeks.org/real-life-applications-of-continuous-probability-distribution/)
* **Probability Functions**

[Probability functions](https://www.geeksforgeeks.org/probability-distribution-function) provide mathematical representations of the probabilities associated with different values of a random variable. Two common types are Probability Mass Functions (PMFs) for discrete variables and [Probability Density Functions](https://www.geeksforgeeks.org/probability-density-function) (PDFs) for continuous variables.

**Statistics Topics**

Some topics of statistics are:

* **Descriptive Statistics**

Descriptive statistics is a branch of statistics focused on summarizing data, presenting it in various forms like graphs or tables. It involves using summary statistics to provide a clear understanding of the data. A descriptive statistic serves as a condensed representation of data. Following are the examples of descriptive statistics given below.

* **Measures of Central Tendency**

[Central Tendency](https://www.geeksforgeeks.org/measures-of-central-tendency) of a set of data is measured by following methods

[**Mean**](https://www.geeksforgeeks.org/what-is-mean/): The average of a set of values. Add up all values and divide by the number of values.

[**Median**](https://www.geeksforgeeks.org/median/): The middle value when data is arranged in order.

[**Mode**](https://www.geeksforgeeks.org/what-is-mode/): The most frequently occurring value in a dataset.

**Learn More**, [Mean, Median and Mode](https://www.geeksforgeeks.org/mean-median-mode)

**Example**: For test scores of 80, 85, 90, 92, and 95, the mean is (80+85+90+92+95)/5 = 88, the median is 90, and the mode is not applicable as there is no repeated value.

* **Measures of Variability**

**Standard Deviation:** Indicates how spread out values are from the mean.

**Variance**: The average of the squared differences from the mean.

Example: In two sets of scores, 70, 75, 80, 85, and 90, and 60, 65, 70, 75, and 80, both have a mean of 80, but the second set has a higher variance, showing more variability.

* **Inferential Statistics**

In practical situations, collecting data from entire populations is often challenging. Descriptive statistics provide a solution by summarizing and organizing available data to offer insights. For instance, calculating the mean (average) and standard deviation from a sample can provide a snapshot of the central tendency and variability in a dataset.

However, when population-scale data collection is impractical, inferential statistics come into play. They involve drawing conclusions about entire populations based on samples. For example, if estimating the mean score of all U.S. high school students on the AP Physics exam is too extensive, inferential statistics enable drawing reliable conclusions from a manageable sample. This approach facilitates informed decision-making even when exhaustive data collection is unfeasible.

* [Covarience and**Correlation**](https://www.geeksforgeeks.org/mathematics-covariance-and-correlation)
* **Data Representations**

[Data representation](https://www.geeksforgeeks.org/what-are-the-different-ways-of-data-representation) involves the presentation of information in a meaningful and understandable manner. In statistics, this is crucial for analyzing and interpreting data effectively. Common methods of data representation include:

* [Graphical Representation](https://www.geeksforgeeks.org/graphical-representation-of-data)
* [Pie Charts](https://www.geeksforgeeks.org/pie-charts)
* [Line Graphs](https://www.geeksforgeeks.org/line-graph)
* [Bar Graphs](https://www.geeksforgeeks.org/bar-graph)
* Scatter Plots
* [Frequency Distribution Tables](https://www.geeksforgeeks.org/frequency-distribution)
* Box-and-Whisker Plots (Boxplots)
* [Dot Plots](https://www.geeksforgeeks.org/dot-plot/)
* [Pictograms](https://www.geeksforgeeks.org/pictographs)
* **Sampling Techniques**

Methods of sampling are used to select a subset of individuals or items from a larger population for the purpose of making inferences about the population. Different sampling techniques are employed based on the nature of the study and the characteristics of the population. Here are some common sampling techniques:

* [Simple Random Sampling](https://www.geeksforgeeks.org/simple-random-sampling/)
* [Stratified Sampling](https://www.geeksforgeeks.org/stratified-sampling-in-machine-learning/#:~:text=Stratified%20sampling%20involves%20dividing%20the,a%20distinct%20category%20or%20label.)
* [Systematic Sampling](https://www.geeksforgeeks.org/systematic-sampling-meaning-types-advantages-and-disadvantages/)
* [Cluster Sampling](https://www.geeksforgeeks.org/cluster-sampling/)
* [Convenience Sampling](https://www.geeksforgeeks.org/non-probability-sampling/#:~:text=1.,it%20convenient%20to%20collect%20information.)
* [Quota Sampling](https://www.geeksforgeeks.org/methods-of-sampling/#:~:text=In%20the%20Quota%20Sampling%20Method,characteristics%20of%20the%20total%20population.)
* [Purposive Sampling](https://www.geeksforgeeks.org/different-types-of-data-sampling-methods-and-techniques/)
* Snowball Sampling

**Probability and Statistics for Engineering Mathematics**

Probability and Statistics form a crucial part of [engineering mathematics](https://www.geeksforgeeks.org/engineering-mathematics-tutorials/), offering a foundation for making informed decisions and solving complex engineering problems. Here’s a brief overview of how these mathematical fields apply to engineering:

**Probability in Engineering**

* **Risk Assessment and Safety Analysis**: Engineers use probability to evaluate the risks associated with different engineering projects or processes, helping to design safer buildings, vehicles, and systems.
* **Quality Control and Reliability Engineering**: Probability models help in assessing the reliability of components and systems, predicting failures, and improving product quality through rigorous testing protocols.
* **Signal Processing**: In electrical and communication engineering, probability is used to analyze and filter signals, dealing with the randomness and noise in data transmission.
* **Decision Making under Uncertainty**: Probability aids in making decisions when outcomes are uncertain, optimizing resources and strategies in situations with incomplete information.

**Statistics in Engineering**

* **Data Analysis and Interpretation**: Engineers collect and analyze data to understand trends, draw conclusions, and support decision-making processes.
* **Experimental Design and Analysis**: Statistical methods are used to design experiments, analyze results, and validate theories or models in fields ranging from material science to environmental engineering.
* **Process and Quality Improvement**: Statistical tools like control charts and design of experiments (DoE) are pivotal in manufacturing and industrial engineering for process optimization and quality enhancement.
* **Predictive Modeling**: Statistics support the creation of models to forecast future events or behaviors, critical in areas such as renewable energy, traffic flow management, and infrastructure development.

**Probability and Statistics – Solved Examples**

**Example 1: Consider the following dataset: [5, 8, 2, 5, 3, 7, 9]. Calculate the mean, median, and mode.**

**Solution:**

*Mean = xˉxˉ*

*xˉxˉ = [5+8+2+5+3+7+9] / 7*

*⇒ 39/7 = 5.579*

*Median:  
The number of values in data set is 7, which is odd n*

*by arranging the values in ascending order [2, 3, 5, 5, 7, 8, 9].*

*The median is the 4th value, which is 5.*

*Mode: The mode is 5, as it appears more frequently than any other number in the dataset.*

**Example 2: Given the dataset [12, 15, 18, 22, 25], calculate the**[**variance and standard deviation**](https://www.geeksforgeeks.org/variance-and-standard-deviation/#:~:text=Variance%20is%20calculated%20by%20taking,is%20expressed%20in%20unit2.)**.**

**Solution:**

*The given data set is [12, 15, 18, 22, 25]*

*Mean = xˉxˉ*

*⇒ xˉxˉ = sum of all values / total number of values*

*⇒ xˉxˉ = (12+15+18+22+25) / 5*

*⇒ 92/5*

*⇒ 18.4*

*Now,*

*Variance = Variance= ∑(Each value−Mean) 2 / Total number of values*

*⇒ σ2 = [(12−18.4)2 + (15−18.4)2 + (18−18.4)2 + (22−18.4)2 + (25−18.4)2 ] / 5*

*⇒ [41.64 + 11.56 + 0.16 + 13.44 + 43.56] /5*

*⇒ 110.36 /5*

*⇒ 22.072*

*We know,*

*Standard deviation = √σ2*

*⇒ √22.072*

*√σ2= 4.69*

**Example 3: In a deck of cards, what is the probability of drawing a red card?**

**Solution:**

*Total number of cards in a deck = 52*

*Total number od Red cards in a deck = 26 (hearts + diamonds)*

*P(Red Card) = 52/26*

*⇒ P(Red Card) = 2/4*

*⇒ P (Red Card) = 1/2 or 0.5 or 50%*

**Practice Problems on Probability and Statistics**

**Problem 1: A bag contains 5 red marbles, 4 blue marbles, and 3 green marbles. What is the probability of randomly selecting a blue marble?**

**Problem 2: A survey is conducted on a sample of 100 people to estimate the average time spent daily on a mobile phone. The sample mean is 2.5 hours with a standard deviation of 1 hour. Calculate a 95% confidence interval for the population mean.**

**Problem 3: A fair six-sided die is rolled. What is the probability of rolling an even number or a number greater than 4?**

**Problem 4: Data Set: [8, 12, 15, 18, 10]. Calculate the variance and standard deviation.**

**Problem 5: Data Set: [10, 15, 12, 18, 15, 22, 20]. Find the mean, median, and mode of the given data set.**

**People Also Read:**

* [*Quartile Formula*](https://www.geeksforgeeks.org/quartile-formula)
* [*Measure of Dispersion*](https://www.geeksforgeeks.org/measure-of-dispersion)
* [*Normal Distribution*](https://www.geeksforgeeks.org/normal-distribution)

**Summary – Probability and Statistics**

Probability and Statistics stand as pivotal tools in understanding the numerical aspects of our world. Probability helps us gauge the chances of occurrences, aiding in predictions like weather forecasting, while statistics dives into data collection and analysis, enabling us to extract meaningful insights from numbers. Together, they empower us to make informed decisions and discern patterns in a sea of information.

This article navigates through various facets of both fields, covering probability’s core principles, such as event types and rules, alongside statistics’ wide array of techniques, from basic data representation to complex inferential analysis. By exploring these domains, we’re equipped with the mathematical understanding to tackle uncertainty and variability, enhancing our grasp of the intricate tapestry of real-world phenomena through a numerical lens.