probability Definitions and Notation

Before discussing the rules of probability, we state the following definitions:

* Two [events](https://stattrek.com/Help/Glossary.aspx?Target=Event) are **mutually exclusive** or **disjoint** if they cannot occur at the same time.
* The probability that Event A occurs, given that Event B has occurred, is called a **conditional probability**. The conditional probability of Event A, given Event B, is denoted by the symbol P(A|B).
* The **complement** of an event is the event not occurring. The probability that Event A will not occur is denoted by P(A').
* The probability that Events A and B *both* occur is the probability of the **intersection** of A and B. The probability of the intersection of Events A and B is denoted by P(A ∩ B). If Events A and B are mutually exclusive, P(A ∩ B) = 0.
* The probability that Events A or B occur is the probability of the **union** of A and B. The probability of the union of Events A and B is denoted by P(A ∪ B) .
* If the occurrence of Event A changes the probability of Event B, then Events A and B are **dependent**. On the other hand, if the occurrence of Event A does not change the probability of Event B, then Events A and B are **independent**.

**Subtraction**

The probability that event A will occur is equal to 1 minus the probability that event A will not occur.

P(A) = 1 - P(A')

**Rule of Multiplication** The probability that Events A and B both occur is equal to the probability that Event A occurs times the probability that Event B occurs, given that A has occurred.

P(A ∩ B) = P(A) P(B|A)

**Rule of Addition** The probability that Event A or Event B occurs is equal to the probability that Event A occurs plus the probability that Event B occurs minus the probability that both Events A and B occur.

P(A ∪ B) = P(A) + P(B) - P(A ∩ B)

Note: Invoking the fact that P(A ∩ B) = P( A )P( B | A ), the Addition Rule can also be expressed as:

P(A ∪ B) = P(A) + P(B) - P(A)P( B | A )

# What is a Random Variable?

When the value of a [variable](https://stattrek.com/Help/Glossary.aspx?Target=Variable) is determined by a chance event, that variable is called a **random variable**.

**Discrete**. Within a range of numbers, discrete variables can take on only certain values.

**Continuous**. Continuous variables, in contrast, can take on any value within a range of values.

# What is a Probability Distribution?

A **probability distribution** is a table or an equation that links each possible value that a [random variable](https://stattrek.com/Help/Glossary.aspx?Target=Random%20variable) can assume with its probability of occurrence.

## Discrete Probability Distributions

The probability distribution of a [discrete](https://stattrek.com/Help/Glossary.aspx?Target=Discrete%20variable) random variable can always be represented by a table.

Continuous Probability Distributions

The probability distribution of a [continuous](https://stattrek.com/Help/Glossary.aspx?Target=Continuous%20variable) random variable is represented by an equation, called the **probability density function** (pdf). All probability density functions satisfy the following conditions:

* The random variable Y is a function of X; that is, y = f(x).
* The value of y is greater than or equal to zero for all values of x.
* The total area under the curve of the function is equal to one.

Independence of Random Variables

If two random variables, X and Y, are **independent**, they satisfy the following conditions.

* P(x|y) = P(x), for all values of X and Y.
* P(x ∩ y) = P(x) \* P(y), for all values of X and Y.

The above conditions are equivalent. If either one is met, the other condition is also met; and X and Y are independent. If either condition is not met, X and Y are **dependent**.

**Note:** If X and Y are independent, then the [correlation](https://stattrek.com/Help/Glossary.aspx?Target=Correlation) between X and Y is equal to zero.

Discrete Variable:

## Binomial Distribution

A **binomial random variable** is the number of successes *x* in *n* repeated trials of a binomial experiment. The [probability distribution](https://stattrek.com/Help/Glossary.aspx?Target=Probability_distribution) of a binomial random variable is called a **binomial distribution**.

## Negative Binomial Distribution

A **negative binomial random variable** is the number *X* of repeated trials to produce *r* successes in a negative binomial experiment. The [probability distribution](https://stattrek.com/Help/Glossary.aspx?Target=Probability_distribution) of a negative binomial random variable is called a **negative binomial distribution**. The negative binomial distribution is also known as the **Pascal distribution**.

Suppose we flip a coin repeatedly and count the number of heads (successes). If we continue flipping the coin until it has landed 2 times on heads, we are conducting a negative binomial experiment. The negative binomial random variable is the number of coin flips required to achieve 2 heads. In this example, the number of coin flips is a random variable that can take on any integer value between 2 and plus infinity. The negative binomial probability distribution for this example is presented below.

|  |  |
| --- | --- |
| **Number of coin flips** | **Probability** |
| 2 | 0.25 |
| 3 | 0.25 |
| 4 | 0.1875 |
| 5 | 0.125 |
| 6 | 0.078125 |
| 7 or more | 0.109375 |

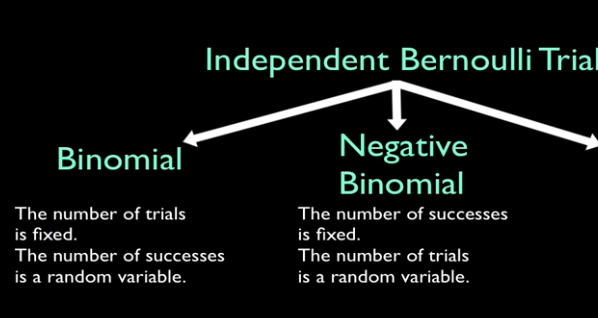
## Geometric Distribution

The **geometric distribution** is a special case of the negative binomial distribution. It deals with the number of trials required for a single success. Thus, the geometric distribution is negative binomial distribution where the number of successes (*r*) is equal to 1.

An example of a geometric distribution would be tossing a coin until it lands on heads. We might ask: What is the probability that the first head occurs on the third flip? That probability is referred to as a **geometric probability** and is denoted by g(*x*; *P*). The formula for geometric probability is given below.

**Geometric Probability Formula**. Suppose a negative binomial experiment consists of *x* trials and results in one success. If the probability of success on an individual trial is *P*, then the geometric probability is:

g(*x*; *P*) = P \* Qx - 1



# The Normal Distribution

The **normal distribution** refers to a family of [continuous probability distributions](https://stattrek.com/Help/Glossary.aspx?Target=Continuous%20probability%20distribution) described by the normal equation.

## The Normal Equation

The normal distribution is defined by the following equation:

**The Normal Equation**. The value of the random variable *Y* is:

Y = { 1/[ σ \* sqrt(2π) ] } \* e-(x - μ)2/2σ2

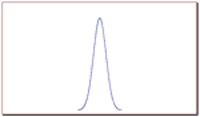
where *X* is a normal random variable, μ is the mean, σ is the standard deviation, π is approximately 3.14159, and *e* is approximately 2.71828.

The random variable *X* in the normal equation is called the **normal random variable**. The normal equation is the [probability density function](https://stattrek.com/Help/Glossary.aspx?Target=Probability%20density%20function) for the normal distribution.

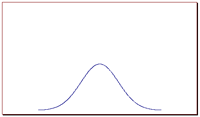
The Normal Curve

The graph of the normal distribution depends on two factors - the mean and the standard deviation. The mean of the distribution determines the location of the center of the graph, and the standard deviation determines the height and width of the graph. All normal distributions look like a symmetric, bell-shaped curve, as shown below.

**Smaller standard deviation**



**Bigger standard deviation**

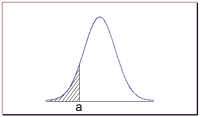


When the standard deviation is small, the curve is tall and narrow; and when the standard deviation is big, the curve is short and wide (see above)

Probability and the Normal Curve

The normal distribution is a continuous probability distribution. This has several implications for probability.

* The total area under the normal curve is equal to 1.
* The probability that a normal random variable *X* equals any particular value is 0.
* The probability that *X* is greater than *a* equals the area under the normal curve bounded by *a* and plus infinity (as indicated by the *non-shaded* area in the figure below).
* The probability that *X* is less than *a* equals the area under the normal curve bounded by *a* and minus infinity (as indicated by the *shaded* area in the figure below).



Additionally, every normal curve (regardless of its mean or standard deviation) conforms to the following "rule".

* About 68% of the area under the curve falls within 1 standard deviation of the mean.
* About 95% of the area under the curve falls within 2 standard deviations of the mean.
* About 99.7% of the area under the curve falls within 3 standard deviations of the mean.

Collectively, these points are known as the **empirical rule** or the **68-95-99.7 rule**. Clearly, given a normal distribution, most outcomes will be within 3 standard deviations of the mean.

# Standard Normal Distribution

The **standard normal distribution** is a special case of the [normal distribution](https://stattrek.com/Help/Glossary.aspx?Target=Normal%20distribution). It is the distribution that occurs when a [normal random variable](https://stattrek.com/Help/Glossary.aspx?Target=Normal%20random%20variable) has a mean of zero and a standard deviation of one.

INFERENTIAL STATISTICS

## Standard Score (aka, z-score)

The normal random variable of a standard normal distribution is called a **standard score** or a **z-score**. Every normal random variable *X* can be transformed into a *z* score via the following equation:

*z* = (*X* - μ) / σ

where *X* is a normal random variable, μ is the mean of *X*, and σ is the standard deviation of *X*.

## Degrees of Freedom

The degrees of freedom refers to the number of independent observations in a set of data.

## Confidence Level

The probability part of a confidence interval is called a **confidence level**. The confidence level describes the likelihood that a particular sampling method will produce a confidence interval that includes the true population parameter.

## Margin of Error

In a confidence interval, the range of values above and below the sample statistic is called the **margin of error**.

## How to Compute the Margin of Error

The margin of error can be defined by either of the following equations.

Margin of error = Critical value x Standard deviation of the statistic  
  
Margin of error = Critical value x Standard error of the statistic

If you know the standard deviation of the statistic, use the first equation to compute the margin of error. Otherwise, use the second equation. Previously, we described [how to compute the standard deviation and standard error](https://stattrek.com/AP-Statistics-4/Standard-Error.aspx?Tutorial=ap).

# What is Hypothesis Testing?

A **statistical hypothesis** is an assumption about a population [parameter](https://stattrek.com/Help/Glossary.aspx?Target=Parameter). This assumption may or may not be true. **Hypothesis testing** refers to the formal procedures used by statisticians to accept or reject statistical hypotheses.

There are two types of statistical hypotheses.

* **Null hypothesis**. The null hypothesis, denoted by Ho, is usually the hypothesis that sample observations result purely from chance.
* **Alternative hypothesis**. The alternative hypothesis, denoted by H1 or Ha, is the hypothesis that sample observations are influenced by some non-random cause.

## One-Tailed and Two-Tailed Tests

A test of a statistical hypothesis, where the region of rejection is on only one side of the [sampling distribution](https://stattrek.com/Help/Glossary.aspx?Target=Sampling_distribution), is called a **one-tailed test**. For example, suppose the null hypothesis states that the mean is less than or equal to 10. The alternative hypothesis would be that the mean is greater than 10. The region of rejection would consist of a range of numbers located on the right side of sampling distribution; that is, a set of numbers greater than 10.

A test of a statistical hypothesis, where the region of rejection is on both sides of the sampling distribution, is called a **two-tailed test**. For example, suppose the null hypothesis states that the mean is equal to 10. The alternative hypothesis would be that the mean is less than 10 or greater than 10. The region of rejection would consist of a range of numbers located on both sides of sampling distribution; that is, the region of rejection would consist partly of numbers that were less than 10 and partly of numbers that were greater than 10.

How to Conduct Hypothesis Tests

All hypothesis tests are conducted the same way. The researcher states a hypothesis to be tested, formulates an analysis plan, analyzes sample data according to the plan, and accepts or rejects the null hypothesis, based on results of the analysis.

* **State the hypotheses.** Every hypothesis test requires the analyst to state a [null hypothesis](https://stattrek.com/Help/Glossary.aspx?Target=Null%20hypothesis) and an [alternative hypothesis](https://stattrek.com/Help/Glossary.aspx?Target=Alternative%20hypothesis). The hypotheses are stated in such a way that they are mutually exclusive. That is, if one is true, the other must be false; and vice versa.
* **Formulate an analysis plan.** The analysis plan describes how to use sample data to accept or reject the null hypothesis. It should specify the following elements.
  + Significance level. Often, researchers choose [significance levels](https://stattrek.com/Help/Glossary.aspx?Target=Significance%20level) equal to 0.01, 0.05, or 0.10; but any value between 0 and 1 can be used.
  + Test method. Typically, the test method involves a test statistic and a [sampling distribution](https://stattrek.com/Help/Glossary.aspx?Target=Sampling%20distribution). Computed from sample data, the test statistic might be a mean score, proportion, difference between means, difference between proportions, z-score, t statistic, chi-square, etc. Given a test statistic and its sampling distribution, a researcher can assess probabilities associated with the test statistic. If the test statistic probability is less than the significance level, the null hypothesis is rejected.
* **Analyze sample data.** Using sample data, perform computations called for in the analysis plan.
  + Test statistic. When the null hypothesis involves a mean or proportion, use either of the following equations to compute the test statistic.

Test statistic = (Statistic - Parameter) / (Standard deviation of statistic)

Test statistic = (Statistic - Parameter) / (Standard error of statistic)

where *Parameter* is the value appearing in the null hypothesis, and *Statistic* is the [point estimate](https://stattrek.com/Help/Glossary.aspx?Target=Point%20estimate) of *Parameter*. As part of the analysis, you may need to compute the standard deviation or standard error of the statistic. Previously, we presented common [formulas for the standard deviation and standard error](https://stattrek.com/AP-Statistics-4/Standard-Error.aspx). When the parameter in the null hypothesis involves categorical data, you may use a chi-square statistic as the test statistic. Instructions for computing a chi-square test statistic are presented in the lesson on the [chi-square goodness of fit test](https://stattrek.com/AP-Statistics-4/Goodness-Of-Fit.aspx).

* + P-value. The P-value is the probability of observing a sample statistic as extreme as the test statistic, assuming the null hypothesis is true.
* **Interpret the results.** If the sample findings are unlikely, given the null hypothesis, the researcher rejects the null hypothesis. Typically, this involves comparing the P-value to the [significance level](https://stattrek.com/Help/Glossary.aspx?Target=Significance%20level), and rejecting the null hypothesis when the P-value is less than the significance level.

# What is the Standard Error?

The standard error is an estimate of the [standard deviation](https://stattrek.com/Help/Glossary.aspx?Target=Standard%20deviation) of a [statistic](https://stattrek.com/Help/Glossary.aspx?Target=Statistic).

## Standard Error of Sample Estimates

Sadly, the values of population parameters are often unknown, making it impossible to compute the standard deviation of a statistic. When this occurs, use the standard error.

The standard error is computed from known sample statistics. The table below shows how to compute the standard error for simple random samples, assuming the population size is at least 20 times larger than the sample size.

|  |  |
| --- | --- |
| **Statistic** | **Standard Error** |
| Sample mean, x | SEx = s / sqrt( n ) |
| Sample proportion, p | SEp = sqrt [ p(1 - p) / n ] |
| Difference between means, x1 - x2 | SEx1-x2 = sqrt [ s21 / n1 + s22 / n2 ] |
| Difference between proportions, p1 - p2 | SEp1-p2 = sqrt [ p1(1-p1) / n1 + p2(1-p2) / n2 ] |

The equations for the standard error are identical to the equations for the standard deviation, except for one thing - the standard error equations use statistics where the standard deviation equations use parameters. Specifically, the standard error equations use p in place of P, and s in place of σ.