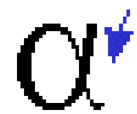
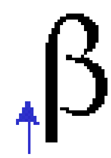
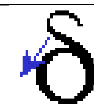
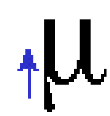
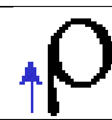
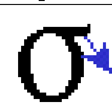
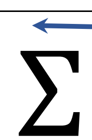
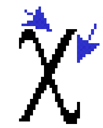
A Brief Survey of Statistical Symbols

Statistics uses many symbols and notation not used in algebra. This document discusses a small collection of the most common ones. We'll begin with some broad categories, and finish with a list of common symbols, their meaning and alternate notations.

Greek Letters

Statistics uses many Greek letters to stand for important concepts or operations. In this section we'll look at 8 of the ones used in a typical elementary statistics course. The third column of the table shows you how to draw the Greek letter. Start at the arrow and follow the direction of the arrow around the letter. Most Greek letters are drawn with a single stroke without overwriting what's been drawn or picking up the pen.

α	The Greek letter **alpha** is used to represent the significance level of a hypothesis test, and $(1-\alpha)$ represents the confidence level of an estimate. The most common value assigned to α is 0.05.	
β	The Greek letter **beta** is used in a couple of different places. It can be used to represent a change of a Type II hypothesis test error, with $(1-\beta)$ being the power of a test (similar to α). It is also used to represent the coefficients in a regressions equation. When it's used in this second way, it usually has subscripts.	
δ	The Greek letter **delta** is used to represent the true difference between two related values. Students typically encounter it in some two-sample hypothesis tests.	
ε	The Greek letter **epsilon** (which can also appear as ϵ) can be used in regression equations as the random error in a measurement.	
μ	The Greek letter **mu** is one of the most commonly used symbols in elementary statistics. It represents the mean of the population.	
ρ	The Greek letter **rho** is used to represent the population correlation of linear regression data.	
σ	The Greek letter **sigma** is one of the most commonly used symbols in statistics. It represents the standard deviation of a population. σ^2 is the variance of a population.	
Σ	The Greek letter **sigma** (this is the capital form) is used to represent the operation addition on a set of numbers. In mathematics, this is sometimes described as a summation, or summation notation.	

χ	The Greek letter **chi** is used in some hypothesis tests. χ^2 is the name of the distribution used to test for independence and goodness-of-fit.	

If you go deeper into statistics, you'll encounter more Greek letters, but this is a good start. There is one other that sometimes comes up that you've seen before: the Greek letter π.

Subscripts

Subscripts are a common notation used in statistics, typically to help reduce the number of symbols needed, or to indicate one element of a set, particularly an ordered set.

x_1 can represent the first element in a list, followed by $x_2, x_3, \ldots etc.$

x_i can represent any element in the list rather than a specific one. You may see this notation in formulas where you need to add up all the elements.

$b_0, \delta_0, etc.$ The zero subscript typically refers to an initial or assumed value for a variable.

$\sigma_{\bar{x}}$ Sometimes subscripts will include other variables. In cases like this is a variable associated with the measurement of another variable. In this example, this is the standard deviation of measurements of the average/mean.

Function Notation

Function notation is a general algebraic rule you may have encountered in algebra. It's used most commonly in probability notation. $P(x)$ is how we notate probabilities, and with the expression inside the parentheses we can say what it's the probability of. Examples: $P(X < 3), P(green)$, etc.

Bars and Hats

A bar over a letter indicates the average or mean of the value. The most common place we see this in statistics is $\bar{x}$, but this notation can be used outside of statistics to mean the same thing. For instance, $\bar{C}$ could represent the average cost.

A hat (or caret) over a letter (^) usually indicates an estimate to distinguish it from a population value. This is mostly commonly used in $\hat{p}$, since p is used for the population instead of π (more on this later). But you can also see it in other situations including $\hat{\mu}, \hat{\sigma}, \hat{\beta}_0$, etc.

Greek vs. Latin

The Greek letters are typically used to represent the population parameters values (that you want to know), and the English or Roman letters are the sample statistic values (that can be measured from collected data).

$\boldsymbol{\mu}$ vs. $\overline{\boldsymbol{x}}$	Mu is the mean of the population while x-bar is the mean of the sample.
$\boldsymbol{\sigma}$ vs. $\boldsymbol{s}$	Sigma is the population standard deviation, while s is the sample standard deviation. These values have different formulas.
$\boldsymbol{\pi}$ vs. $\boldsymbol{p}$ vs. $\widehat{\boldsymbol{p}}$	Since the letter pi is used for the ratio of a radius to the circumference of a circle, it's rarely used to represent the population proportion. If π is used, then p may be used for the sample. When π is not used, then p is used for the population proportion and then $\hat{p}$ is the sample proportion. The latter situation is perhaps the most common in elementary statistics.
$\boldsymbol{\rho}$ vs. $\boldsymbol{r}$	Rho is the population correlation between two variables, and r is the sample correlation.
$\boldsymbol{\beta_1}$ vs. $\boldsymbol{b_1}$	In this usage, β_1 is the coefficient of the linear term (the slope) in a linear regression problem, representing the slope of the true relationship in the population. b_1 is the same coefficient estimated from a sample. A similar relationship exists for other regression coefficients.

Table of common symbols

Symbol	Definition and Alternates
$P(A)$	Probability of event A
$\cap$	Intersection of sets; $A \cap B$ means things in BOTH A and in B
$\cup$	Union of sets; $A \cup B$ means things in either A or B
$-$	Subtraction in set notation; $A - B$ means all the elements in A that are not also in B, can stand for the sequence of operations $A \cap B^c$
$f(x)$	Function notation; in statistics can stand for a probability density function
A^c	Complementation; all the elements not in A. There are various notations used that mean the same thing, including $not\ A$, $\sim A$, $\bar{A}$, A', $A^{\neg}$, etc.
μ	The population mean (lower case Greek mu)
σ	The population standard deviation (lower case Greek sigma)
σ^2	The population variance
$\bar{x}$	The sample mean, read "x-bar"
s	The sample standard deviation
s^2	The sample variance
Σ	Capital Sigma; Summation notation; it means add up according the to the formula. Notation under the Σ will tell you where to start, and above it, where to end. If none is given, add up everything that applies.
ρ	Population correlation coefficient

r	The sample correlation coefficient
ϵ	The Greek letter epsilon can stand for a "small amount", and error term in an equation, or in set notation, means "an element of", so $x \in A$ means "x is an element of the set A". In the usage of a small amount or error, can also be written as ε.
$\tilde{x}$	The median. Sometimes notated as Q_2 or M
Q_1	The first Quartile, or 25th percentile
Q_3	The third Quartile, or 75th percentile
z	Standard score of an observation
z_α	Critical value of z associated with α% above the given value of z in a standard normal distribution.
!	Factorial; multiply all the integers from the specified number down to 1; $7! = 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1$
nPr	Permutations; $nPr = \frac{n!}{(n-r)!}$; you may also see other notations like $P(n,r)$ or with subscripts and superscripts
nCr	Combinations; $nCr = \frac{n!}{r!(n-r)!}$; you may also see other notations like $C(n,r)$ or with subscripts and superscripts
$\binom{n}{r}$	Combinations = nCr, alternative combinations notation common in statistics
IQR	Interquartile Range; $Q_3 - Q_1$
x^*	Sometimes used to specify a value of the independent variable where a prediction is being made.
X	Capital letters are used to denote random variables; their values are denoted by lower case letters
x_i	The i^{th} measurement of x
y_i	The i^{th} measurement of y, corresponding to x_i
$\hat{y}_i$	The expected or mean value of y_i obtained from a regression equation and not an observation
p	The population proportion; sometimes this will be denoted by π
$\hat{p}$	The estimated or sample proportion
χ^2	χ^2 is used to describe a distribution used in independence tests, and goodness-of-fit tests
$\beta_0 + \beta_1 x$	The betas here are used to describe the coefficients in a regression equation; in a linear regression equation like this, there are only two coefficients, but in polynomial regression or multiple regression equations there can be many coefficients. The $\beta_i s$ are estimated with b_i from a sample.
m	We use "m" for mode, but it does not have a standard symbol. Not to be confused with the slope of a line, which in statistics is β_1.
$t_{\alpha,\nu}$	The critical value for the Student-t distribution needs two subscripts, one for the probability (like z_α), but also for the degrees of freedom
ν	This is not a "V", but a Greek letter lower case "nu". This can be used for the degrees of freedom, or you may see the abbreviation df used instead.

q	$= 1 - p$; since the expression $1 - p$ is used so often in statistics (the complement of the population proportion), it has its own symbol. Likewise, $\hat{q} = 1 - \hat{p}$.
n	The number of elements in a sample.
r^2, R^2	Coefficient of determination; lower case is used for linear or intrinsically linear regressions, while upper case is used in polynomial and multiple regression.
$\emptyset$	Empty set; Can also be notated as $\{\quad\}$.
α	The significance level of a hypothesis test; $1 - \alpha$ is the confidence level
β	The probability of a Type II error in hypothesis testing; $1 - \beta$ is the power of a hypothesis test
H_0	The null hypothesis; notation used to indicate default assumption before conducting test
H_a	The alternative hypothesis; the conclusion you are trying to prove; may also be notated as H_A or H_1
N	The total number of elements in the population.
δ	Sometimes used for the difference between two dependent samples.