0: Notation and Symbols Used in Statistics

Statistics (and mathematics) uses symbols to make things easier and clearer. Symbols help express complex ideas quickly and efficiently, allowing us to communicate universally without language barriers. They ensure precision and avoid ambiguity, making understanding and working with mathematical concepts easier. Using symbols, we can manage complexity, represent abstract ideas, and recognize patterns more effectively. This way, math becomes a powerful tool for exploring and understanding the world around us. By the end of this section, readers will be equipped with the notation skills necessary to navigate through statistical literature and confidently perform their own analyses.

Note: Population versus Sample

Formal definitions for **population** and **sample** will be provided in 1.2: Definitions of Statistics, Probability, and Key Terms.

In statistics, a **population** refers to the entire group of individuals or items that we are interested in studying. It includes all possible subjects that fit the criteria of the research.

For example, we are studying the heights of adult women in Fresno. The 2024 population of Fresno, California was 641,528 people of which 190,584 are adult women living in Fresno. Therefore the population for our study would be 190,584 adult women living in Fresno.

A **sample**, on the other hand, is a subset of the population selected for the actual study. It represents a smaller group chosen from the population, ideally in a way that accurately reflects the larger group.

Continuing with the same example, a sample might include a smaller group of 400 adult women from different parts of Fresno.

Symbols: Greek and Latin Alphabet

Statistics uses Greek and Latin letters for clarity, tradition, and effectively distinguishing different types of data. Typically, Greek letters describe population information, while Latin (or Latin-looking) letters represent sample information.

Latin Alphabet

Aa, Bb, Cc, Dd, Ee, Ff, Gg, Hh, Ii, Jj, Kk, Ll, Mm, Nn, Oo, Pp, Qq, Rr, Ss, Tt, Uu, Vv, Ww, Xx, Yy, Zz

Greek Alphabet

The **highlighted** letters are letters we use in this book.

Uppercase, Lowercase	Name	Uppercase, Lowercase	Name	Uppercase, Lowercase	Name
A, $\boldsymbol{\alpha}$	alpha	I, ι	iota	P, $\boldsymbol{\rho}$	rho
B, $\boldsymbol{\beta}$	beta	K, κ	kappa	Σ, $\boldsymbol{\sigma}$	sigma
Γ, γ	gamma	Λ, $\boldsymbol{\lambda}$	lambda	T, τ	tau
Δ, δ	delta	M, $\boldsymbol{\mu}$	mu	Υ, υ	upsilon
E, ϵ	epsilon	N, ν	nu	Φ, ϕ	phi
Z, ζ	zeta	Ξ, ξ	xi	X, $\boldsymbol{\chi}$	chi
H, η	eta	O, o	omicron	Ψ, ψ	psi
Θ, θ	theta	Π, $\boldsymbol{\pi}$	pi	Ω, ω	omega

Symbols: Uppercase and Lowercase Lettering

Lower and uppercase letters are used to distinguish between different types of quantities and concepts. This notation helps provide clarity and consistency in statistical communication.

- **Lowercase Letters**:
  - Typically used for sample statistics, individual data points, and variables.
  - Examples: x - a single data point, n - sample size, $\bar{x}$ - sample standard deviation, μ - population mean, σ - population standard deviation
- **Uppercase Letters**:
  - Typically used for population parameters, random variables, and distributions.
  - Examples: X - population random variable, N - Normal Distribution, S - might represent Sample Space (the set of all possible outcomes of an experiment).

Definition 0.2 Distribution Notation

Distribution notation in mathematics and statistics is used to describe how values of a random variable are spread or distributed. This notation conveys information about the probability distribution that a random variable follows, allowing us to understand its behavior and make predictions based on it.

Common Distribution Notations

- Discrete Distributions:
  - **Binomial Distribution**: $X \sim B(n, p)$
    - **Meaning**: Random variable X follows a binomial distribution with n trials and probability of success p in each trial.
  - **Poisson Distribution**: $X \sim P(\mu)$
    - **Meaning**: Random variable X follows a Poisson distribution with meanμ.
- Continuous Distributions:
  - **Normal Distribution**: $X\ N(\mu, \sigma)$
    - **Meaning**: Random variable X follows a normal distribution with mean, μ, and standard deviation, σ.
  - **Exponential Distribution**: $X \sim Exp(\lambda)$
    - **Meaning**: Random variable X follows an exponential distribution with rate parameter λ.

Definition 0.1 Subscript (Indexing)

Index or Subscript: An index or subscript on variables is a notation used to distinguish between multiple related variables, typically in contexts involving sequences, arrays, or matrices. The notation of the index (subscript) is a small number, letter, or symbol written slightly below and to the right of a variable.

Example: $x_1, x_2, x_3, \ldots, x_n$ where x_i represents the i^th element in a sequence.

Example: If we want to add up $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}$, we can use summation notation where we start our indexing at 1 and go up to n = 10.

$$\sum_{n=1}^{10} x_n = x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 + x_9 + x_{10}$$

Symbols: Summation Notation

Given a sequence $\{a_n\}_{n=k}^{\infty}$ and numbers m and p satisfying $k \leq m \leq p$, the summation from m to p of the sequence $\{a_n\}$ is written

$$\sum_{n=m}^{p} a_n = a_m + a_{m+1} + \ldots + a_p$$

The variable n is called the **index of summation**. The number m is called the **lower limit of summation** while the number p is called the **upper limit of summation**.

In English, Definition 0.1 is simply defining a short-hand notation for adding up the terms of the sequence $\{a_n\}_{n=k}^{\infty}$ from a_m through a_p. The symbol Σ is the capital Greek letter sigma and is shorthand for 'sum'. The lower and upper limits of the summation tells us which term to start with and which term to end with, respectively. For example, using the sequence $a_n = 2n - 1$ for $n \geq 1$, we can write the sum $a_3 + a_4 + a_5 + a_6$ as

$$\begin{aligned} \sum_{n=3}^{6}(2n-1) &= (2(3)-1)+(2(4)-1)+(2(5)-1)+(2(6)-1) \\ &= 5+7+9+11 \\ &= 32 \end{aligned}$$

The index variable is considered a 'dummy variable' in the sense that it may be changed to any letter without affecting the value of the summation. For instance,

$$\sum_{n=3}^{6}(2n-1)=\sum_{k=3}^{6}(2k-1)=\sum_{j=3}^{6}(2j-1)$$

One place you may encounter summation notation is in mathematical definitions. For example, summation notation allows us to define polynomials as functions of the form

$$f(x)=\sum_{k=0}^{n} a_k x^k$$

for real numbers a_k, $k = 0, 1, \ldots n$. The reader is invited to compare this with what is given in Definition 3.1. Summation notation is particularly useful when talking about matrix operations. For example, we can write the product of the ith row R_i of a matrix $A = [a_{ij}]_{m\times n}$ and the j^{th} column C_j of a matrix $B = [b_{ij}]_{n\times r}$ as

$$Ri \cdot Cj=\sum_{k=1}^{n} a_{ik} b_{kj}$$