

# Subscript and Summation Notation

James H. Steiger

## Single Subscript Notation

- # Most of the calculations we perform in statistics are repetitive operations on lists of numbers. For example, we compute the sum of a set of numbers, or the sum of the squares of the numbers, in many statistical formulas. We need an efficient notation for talking about such operations in the abstract.

## Single Subscript Notation

List Name  $X$   
Subscript  $i$

## Single Subscript Notation

- # The symbol  $X$  is the “list name,” or the name of the variable represented by the numbers on the list. The symbol  $i$  is a “subscript,” or “position indicator.” It indicates which number in the list, starting from the top, you are referring to.

$X_3$ 

## Single Subscript Notation

---

$X$
1
2
12
3
14

←  $X_3$

 $X_3$ 

## Single Subscript Notation

---

- Single subscript notation extends naturally to a situation where there are two or more lists. For example suppose a course has 4 students, and they take two exams. The first exam could be given the variable name  $X$ , the second  $Y$ , as in the table below. Chow's score on the second exam is observation  $Y_2$

 $X_3$ 

## Single Subscript Notation

---

<i>Student</i>	$X$	$Y$
Smith	87	85
Chow	65	66
Benedetti	83	90
Abdul	92	97

## Double Subscript Notation

---

- Using different variable names to stand for each list works well when there are only a few lists, but it can be awkward for at least two reasons.
  - In some cases the number of lists can become large. This arises quite frequently in some branches of psychology.
  - When general theoretical results are being developed, we often wish to express the notion of some operation being performed "over all of the lists." It is difficult to express such ideas efficiently when each list is represented by a different letter, and the list of letters is in principle unlimited in size.

## Double Subscript Notation

$$X_{ij}$$

## Double Subscript Notation

- # The first subscript refers to the row that the particular value is in, the second subscript refers to the column.

$$\begin{array}{ccc} X_{11} & X_{12} & X_{13} \\ X_{21} & X_{22} & X_{23} \\ X_{31} & X_{32} & X_{33} \end{array}$$

## Double Subscript Notation

- # Test your understanding by identifying  $X_{32}$  in the table below.

$$\begin{array}{ccc} 1 & 43 & 13 \\ 23 & 42 & 33 \\ 12 & 11 & 23 \end{array}$$

## Single Summation Notation

- # Many statistical formulas involve repetitive summing operations. Consequently, we need a general notation for expressing such operations.
- # We shall begin with some simple examples, and work through to some that are more complex and challenging.

## Single Summation Notation

- Many summation expressions involve just a single summation operator. They have the following general form

$$\sum_{i=1}^N X_i$$

Annotations:  
-  $N$ : stop value  
-  $X_i$ : summation index  
-  $i=1$ : start value

## Rules of Summation Evaluation

- The summation operator governs everything to its right, up to a natural break point in the expression.
- Begin by setting the *summation index* equal to the start value. Then evaluate the algebraic expression governed by the summation sign.
- Increase the value of the *index* by 1. Evaluate the expression governed by the summation sign again, and add the result to the previous value.
- Keep repeating step 3 until the expression has been evaluated and added for the *stop value*. At that point the evaluation is complete, and you stop.

## Evaluating a Simple Summation Expression

- Suppose our list has just 5 numbers, and they are 1,3,2,5,6. Evaluate

$$\sum_{i=1}^5 X_i^2$$

Answer:

$$1^2 + 3^2 + 2^2 + 5^2 + 6^2 = 75$$

## Evaluating a Simple Summation Expression

- Order of evaluation can be crucial. Suppose our list is still 1,3,2,5,6. Evaluate

$$\left( \sum_{i=1}^5 X_i \right)^2$$

Answer:

$$(1 + 3 + 2 + 5 + 6)^2 = 17^2 = 289$$

## The Algebra of Summations

---

- # Many facts about the way lists of numbers behave can be derived using some basic rules of summation algebra. These rules are simple yet powerful.
  - The first constant rule
  - The second constant rule
  - The distributive rule

## The First Constant Rule

---

- # The first rule is based on a fact that you first learned when you were around 8 years old: multiplication is simply repeated addition.
- # That is, to compute 3 times 5, you compute  $5+5+5$ .
- # Another way of viewing this fact is that, if you add a constant a certain number of times, you have multiplied the constant by the number of times it was added.

## The First Constant Rule

---

- # Symbolically, we can express the rule as:

$$\sum_{i=x}^y a = (y - x + 1)a$$

## The First Constant Rule (Simplified Version)

---

- # Symbolically, we can express the rule as:

$$\sum_{i=1}^N a = Na$$

## The First Constant Rule (Application Note)

---

- # The symbol  $a$  refers to *any expression, no matter how complicated*, that does not vary as a function of  $i$ , the summation index! Do not be misled by the form in which the rule is expressed.

## The First Constant Rule (Example)

---

- # Evaluate the following: (C.P.)

$$\sum_{i=2}^3 2$$

## The First Constant Rule (Example)

---

- # Simplify the following: (C.P.)

$$\sum_{i=1}^N (2X_j - 5)^2$$

## The Second Constant Rule

---

- # The second rule of summation algebra, like the first, derives from a principle we learned very early in our educational careers. When we were first learning algebra, we discovered that a common multiple could be factored out of additive expressions. For example,

$$2x + 2y = 2(x + y)$$

## The Second Constant Rule

- ‡ The rule states that

$$\sum_{i=1}^N aX_i = a \sum_{i=1}^N X_i$$

- ‡ Again, the rule appears to be saying less than it actually is. At first glance, it appears to be a rule about multiplication. You can move a factorable constant outside of a summation operator. However, the term  $a$  could also stand for a fraction, and so the rule also applies to factorable divisors in the summation expression.

## The Second Constant Rule (Examples)

- ‡ Apply the Second Constant Rule to the following:

$$\sum_{i=1}^N 6yX_i$$

$$\sum_{i=1}^N \frac{X_i}{2}$$

## The Distributive Rule of Summation Algebra

- ‡ The third rule of summation algebra relates to another fact that we learned early in our mathematics education --- when numbers are added or subtracted, the ordering of addition and/or subtraction doesn't matter. For example

$$(1 + 2) + (3 + 4) = (1 + 2 + 3 + 4)$$

## The Distributive Rule of Summation Algebra

- ‡ So, in summation notation, we have

$$\sum_{i=1}^N (X_i + Y_i) = \sum_{i=1}^N X_i + \sum_{i=1}^N Y_i$$

Since either term could be negative, we also have

$$\sum_{i=1}^N (X_i - Y_i) = \sum_{i=1}^N X_i - \sum_{i=1}^N Y_i$$

## Definition: The Sample Mean and Deviation Scores

---

- # The *sample mean* of  $N$  scores  $X_i$  is defined as their arithmetic average,

$$\bar{X}_{\bullet} = \frac{1}{N} \sum_{i=1}^N X_i$$

The original scores are called *raw scores*.

The *deviation scores* corresponding to the raw scores are defined as

$$dx_i = X_i - \bar{X}_{\bullet}$$

---