

1 Summation Notation

1.1 Introduction

Suppose you were interested in taking the mean (or average) of a list of values, for example: 3, 5, 2, 7, 8. You would first add the values together, and then divide by five, which is the number of quantities. One way to write this would be:

$$\frac{3 + 5 + 2 + 7 + 8}{5} = \frac{25}{5} = 5$$

Now, suppose you had a very long list of values. It would be useful to have some notation so that you could write an expression for the average without having to write out all of the individual values. We can do that using subscripts to identify the individual values and “summation notation.”

1.2 Using subscripts and Σ -notation

Continuing with our example above, let’s call the number of quantities n , so that in our case $n = 5$. Next, call the first value x_1 , the second value x_2 , and so on, so that with our list of values we have $x_1 = 3, x_2 = 5, x_3 = 2, x_4 = 7$, and $x_5 = 8$. We can call the last number in our list x_n , since there are five numbers in our list and $n = 5$. So $x_n = x_5 = 8$, in this example.

Subscripts give us an easy way of referring to any single quantity in the list. For instance, x_3 is the short hand notation for “the third element in the list”. Now, if we used our subscript notation to write out the expression for the numerator in the mean, we would have $x_1 + x_2 + x_3 + x_4 + x_5$, or we might write this as $x_1 + x_2 + \cdots + x_5$, but this is still a bit long to write out. A shorter form of notation for this same expression using the summation sign (Σ) would be:

$$\sum_{i=1}^n x_i .$$

This can be read as “the sum of all x_i for $i = 1$ through $i = n$.” In our case, it would be the sum of the numbers from x_1 through $x_n = x_5$, so the sum of all of the values in our list.

1.2.1 Example

Let $n = 3$, with $x_1 = 7, x_2 = -8$, and $x_3 = 13$.

$$\text{Then, } \sum_{i=1}^n x_i = \sum_{i=1}^3 x_i = x_1 + x_2 + x_3 = 7 + (-8) + 13 = 12.$$

To calculate the average of these three values we would take the sum of all three numbers divided by three. If we use summation notation, we would write:

$$\text{average} = \frac{1}{3} \sum_{i=1}^3 x_i = \frac{x_1 + x_2 + x_3}{3} = \frac{7 + (-8) + 13}{3} = 12/3 = 4.$$

1.3 Common statistical formulas that use summations

You probably recognize that in our examples so far, we have been working with the formula for the sample mean, which can be written like this using summation notation:

$$\text{sample mean} = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i. \quad (1.3.1)$$

Compare this to the formula we used in Example 1.2.1. This general formula will work for a list of any length. The number n represents the number of quantities in the list, and the terms x_i represent the actual values.

The formula for the sample standard deviation can also be written using summation notation:

$$\text{sample standard deviation} = s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}} \quad (1.3.2)$$

Let's look at an example.

1.3.1 Example

Suppose we wanted to calculate the sample standard deviation for our list of values from Example 1.2.1. Let's think about what formula (1.3.2) is telling us. We can read the summation part of the formula as "the sum of the terms $(x_i - \bar{x})^2$ for $i = 1$ to $i = n$." So we can think of *expanding* out the terms in the numerator like this:

$$\sum_{i=1}^n (x_i - \bar{x})^2 = (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2$$

If we substitute that into our formula for the sample standard deviation we get this:

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}} = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n - 1}}$$

In Example 1.2.1 we found that the average of our values is $\bar{x} = 4$, and there are $n = 3$ numbers in the list. If we substitute $\bar{x} = 4$ and $n = 3$ into our expanded formula for the sample standard deviation we get:

$$s = \sqrt{\frac{(x_1 - 4)^2 + (x_2 - 4)^2 + (x_3 - 4)^2}{3 - 1}}$$

and then we can plug in our specific values, $x_1 = 7$, $x_2 = -8$ and $x_3 = 13$, to get:

$$\begin{aligned}
 s &= \sqrt{\frac{(7-4)^2 + ((-8)-4)^2 + (13-4)^2}{2}} \\
 &= \sqrt{\frac{(3)^2 + (-12)^2 + (9)^2}{2}} \\
 &= \sqrt{\frac{9 + 144 + 81}{2}} \\
 &= \sqrt{\frac{234}{2}} = \sqrt{117} = 10.82
 \end{aligned}$$

So the sample standard deviation of the list of values in Example 1.2.1 is $s = 10.82$.

1.4 Useful rules for working with summations

1.4.1 A “Multiplicative” Rule

$$\boxed{\sum_{i=1}^n (a \cdot x_i) = a \cdot \sum_{i=1}^n x_i}$$

Why is this true? Let’s work an example with specific values. Suppose we have $x_1 = 3$, $x_2 = 5$, and $x_3 = -1$, and we want to calculate $\sum_{i=1}^3 (5 \cdot x_i)$. That would be the sum of the three terms that each look like this: $(5 \cdot x_i)$ for $i = 1$ to $i = 3$.

$$\begin{aligned}
 \sum_{i=1}^3 (5 \cdot x_i) &= (5 \cdot x_1) + (5 \cdot x_2) + (5 \cdot x_3) = (5 \cdot 3) + (5 \cdot 5) + (5 \cdot (-1)) \\
 &= 5 \cdot (3 + 5 + (-1)) = 5 \cdot (x_1 + x_2 + x_3) \\
 &= 5 \cdot \sum_{i=1}^3 x_i
 \end{aligned}$$

This is quite handy. Whenever your terms in a summation are multiplied by a constant, we can pull that constant outside of the summation.

1.4.2 An “Additive” Rule

$$\boxed{\sum_{i=1}^n (x_i + y_i) = \sum_{i=1}^n x_i + \sum_{i=1}^n y_i}$$

Whenever you have an addition or subtraction sign in the summation, you can break the equation apart into separate summations.

To see that this is true, suppose you have two lists of numbers that are both the same length. The first list has $x_1 = 3$, $x_2 = 5$, $x_3 = -1$, and the second list of numbers has $y_1 = 7$, $y_2 = 4$, and $y_3 = 2$. Then,

$$\sum_{i=1}^3 (x_i + y_i) = (x_1 + y_1) + (x_2 + y_2) + (x_3 + y_3) = (3 + 7) + (5 + 4) + ((-1) + 2).$$

Since we are just adding a list of numbers together, we know from basic algebra that we can change the order of addition without changing the result (the commutative property of addition). So, we could rewrite the last sum above as $(3 + 5 + (-1)) + (7 + 4 + 2)$, which is just the sum of all of the numbers in the first list plus the sum of all of the numbers in the second list. If we write this all out carefully using summation notation we get:

$$\sum_{i=1}^3 (x_i + y_i) = (3+7) + (5+4) + ((-1)+2) = (3+5+(-1)) + (7+4+2) = \sum_{i=1}^3 x_i + \sum_{i=1}^3 y_i.$$

The additive rule really only works for “additive functions.” That is, the rule works for functions of the form $f(x_i) = g(x_i) + h(x_i) + p(x_i)$. In fact, a more general form for the additive rule is this:

$$\sum_{i=1}^n [g(x_i) + h(x_i) + p(x_i)] = \sum_{i=1}^n g(x_i) + \sum_{i=1}^n h(x_i) + \sum_{i=1}^n p(x_i)$$

Caution:

$$\sum_{i=1}^n (a + x_i)^2 \neq \sum_{i=1}^n a^2 + \sum_{i=1}^n x_i^2.$$

However, we can still use the additive rule here as follows:

$$\sum_{i=1}^n (a + x_i)^2 = \sum_{i=1}^n (a^2 + 2ax_i + x_i^2) = \sum_{i=1}^n a^2 + \sum_{i=1}^n 2ax_i + \sum_{i=1}^n x_i^2.$$

1.4.3 A “Constant” Rule

$$\sum_{i=1}^n a = n \cdot a$$

This rule says that if you have a summation of a *constant* (a single value that does not change with the index i), then the result will just be that number times n . Why is this true? Let’s use $a = 7$ and $n = 5$. Then we would get:

$$\sum_{i=1}^n a = \sum_{i=1}^5 7 = 7 + 7 + 7 + 7 + 7 = 5 \cdot 7 = 35.$$

1.4.4 A “Derivative” Rule

$$\frac{d}{du} \left[\sum_{i=1}^n f_i(u) \right] = \sum_{i=1}^n \left[\frac{d}{du} f_i(u) \right] \quad (\text{when } n < \infty)$$

In this course, we will study *linear models* and the theory of *least squares* estimates. It will be important to know how to take the derivative of a function of this form (with a summation). As long as you have a finite number of terms in your summation (for this class, this will always be the case), you can always move a derivative operation into the summation. Here is a short example:

Suppose we wanted to show that

$$\frac{d}{du} \left[\sum_{i=1}^n (ux_i^2 + u^2x_i) \right] = \sum_{i=1}^n (x_i^2 + 2ux_i).$$

Using our new derivative rule along with this fact from calculus: $\frac{d}{du}(ux_i^2 + u^2x_i) = (x_i^2 + 2ux_i)$,

$$\frac{d}{du} \left[\sum_{i=1}^n (ux_i^2 + u^2x_i) \right] = \sum_{i=1}^n \frac{d}{du} (ux_i^2 + u^2x_i) = \sum_{i=1}^n (x_i^2 + 2ux_i)$$

and by our additive and multiplicative rules, we have that:

$$\sum_{i=1}^n (x_i^2 + 2ux_i) = \sum_{i=1}^n x_i^2 + \sum_{i=1}^n 2ux_i = \sum_{i=1}^n x_i^2 + 2u \sum_{i=1}^n x_i,$$

and this seems to be all of the algebraic simplification that we can perform.

However, this is a statistics course and if x_1, x_2, \dots, x_n are a random sample of observations from a population, then we can show (try this own your own) that

$$\sum_{i=1}^n x_i = n\bar{x} \quad \text{and} \quad \sum_{i=1}^n x_i^2 = (n-1)s^2 + n\bar{x}^2$$

by equations (1.3.1) and (1.3.2) and a bit of algebra using summation notation. Finally,

$$\frac{d}{du} \left[\sum_{i=1}^n (ux_i^2 + u^2x_i) \right] = \sum_{i=1}^n x_i^2 + 2u \sum_{i=1}^n x_i = (n-1)s^2 + n\bar{x}^2 + 2un\bar{x} = (n-1)s^2 + n\bar{x}(\bar{x} + 2u),$$

and this seems to be all of the “statistical” simplification that we can perform.

1.5 Practice Problems

This last section will present you with five problems that utilize the rules presented above. The problems are intended to show the most complicated level of mathematics involving summations that you will encounter in this course. You should try to master the steps in these problems in order to avoid having difficulty with following “proofs” that make use of summation notation later in the course.

In what follows, a, b, t, x, y and u are real-valued numbers (in \mathbb{R}), while n and m are integers (in \mathbb{N}). The answers appear on the pages following the questions.

1.5.1 Practice Problem Questions

1. Show that

$$\sum_{i=1}^n (ax_i - b)^2 = a^2 \sum_{i=1}^n x_i^2 - 2ab \sum_{i=1}^n x_i + nb^2$$

is true by expanding the sum on the left-hand side into its individual terms, then rearranging and collecting terms together appropriately to obtain the sum on the right hand side.

2. Provide a counterexample to show that it is generally **not true** that

$$\sum_{i=1}^n x_i^2 = \left[\sum_{i=1}^n x_i \right] \left[\sum_{i=1}^n x_i \right].$$

That is, for something like $n = 3$, choose some actual values for x_1, x_2 and x_3 such that the two sides of the equation are not equal.

3. Find the derivative of $f(a) = \sum_{i=1}^n (ax_i - b)^2$ with respect to a .

4. Minimize $f(a) = \sum_{i=1}^n (ax_i - b)^2$ with respect to a .

Be sure to confirm that any extrema that you find is indeed a minimum (not a maximum).

5. Show that

$$\sum_{i=1}^n \sum_{j=1}^m (x_i - a)(y_j - b) = \left[\sum_{i=1}^n x_i \right] \left[\sum_{j=1}^m y_j \right] - mb \sum_{i=1}^n x_i - na \sum_{j=1}^m y_j + (mn)(ab).$$

If this seems complicated, first use $n = 3$ and $m = 2$ to limit the number of individual terms that you need to write down in order to get comfortable working with double sums, then go back and work through the general case for any n and m .

See next pages for answers to these questions.

1.5.2 Practice Problem Answers

1.

$$\begin{aligned}\sum_{i=1}^n (ax_i - b)^2 &= \sum_{i=1}^n (a^2x_i^2 - 2abx_i + b^2) \\ &= \sum_{i=1}^n a^2x_i^2 + \sum_{i=1}^n (-2)abx_i + \sum_{i=1}^n b^2 \quad (\text{by the additive rule}) \\ &= a^2 \sum_{i=1}^n x_i^2 - 2ab \sum_{i=1}^n x_i + nb^2 \quad (\text{by the multiplicative and constant rules})\end{aligned}$$

2. Suppose $n = 3$ and $x_1 = 2, x_2 = 3, x_3 = 1$. Then,

$$\sum_{i=1}^n x_i^2 = 2^2 + 3^2 + 1^2 = 4 + 6 + 2 = 12, \quad \text{while} \quad \left[\sum_{i=1}^n x_i \right] \left[\sum_{i=1}^n x_i \right] = (2+3+1)(2+3+1) = 6^2 = 36.$$

3.

$$\begin{aligned}\frac{d}{da} f(a) &= \frac{d}{da} \sum_{i=1}^n (ax_i - b)^2 \\ &= \sum_{i=1}^n \frac{d}{da} (ax_i - b)^2 \quad (\text{by the derivative rule}) \\ &= \sum_{i=1}^n 2(ax_i - b)(x_i) \quad (\text{don't forget the chain rule for derivatives!}) \\ &= 2a \sum_{i=1}^n x_i^2 - 2b \sum_{i=1}^n x_i \quad (\text{by the multiplicative and additive rules})\end{aligned}$$

4. Set the derivative in the previous question equal to zero and solve for a :

$$\begin{aligned}2a \sum_{i=1}^n x_i^2 - 2b \sum_{i=1}^n x_i &= 0 \\ 2a \sum_{i=1}^n x_i^2 &= 2b \sum_{i=1}^n x_i \\ a &= \frac{b \sum_{i=1}^n x_i}{\sum_{i=1}^n x_i^2}\end{aligned}$$

Be very careful. It is **not true** that $a = \frac{b}{\sum_{i=1}^n x_i}$. See question 2 for the reason.

To check that this is indeed a minimum and not a maximum, we will check that the sign of the second derivative of $f(a)$ is positive at the extreme value.

$$\frac{d^2}{da^2} f(a) = \frac{d}{da} \left[\frac{d}{da} f(a) \right] = \frac{d}{da} \left[2a \sum_{i=1}^n x_i^2 - 2b \sum_{i=1}^n x_i \right] = 2 \sum_{i=1}^n x_i^2 > 0$$

for any $a \in \mathbb{R}$. Of course, $\sum_{i=1}^n x_i^2$ could be $= 0$, but this would be a very unlikely (and uninteresting) set of data with each and every value $x_i = 0$ and the function would be constant not involving a : $f(a) = nb^2$. Thus, the extreme value we have found is indeed a minimum.

5. Let's try this out with $n = 3$ and $m = 2$ first, as suggested.

$$\begin{aligned} & \sum_{i=1}^n \sum_{j=1}^m (x_i - a)(y_j - b) \\ &= \sum_{i=1}^3 \sum_{j=1}^2 (x_i - a)(y_j - b) \\ &= \sum_{i=1}^3 \sum_{j=1}^2 (x_i y_j - b x_i - a y_j + ab) \\ &= \sum_{i=1}^3 [(x_i y_1 - b x_i - a y_1 + ab) + (x_i y_2 - b x_i - a y_2 + ab)] \quad (\text{by summing over } j = 1, 2) \\ &= \sum_{i=1}^3 [(x_i y_1 + x_i y_2) + (-b x_i - b x_i) + (-a y_1 - a y_2) + (ab + ab)] \quad (\text{by collecting up like terms}) \\ &= \sum_{i=1}^3 (x_i y_1 + x_i y_2) + \sum_{i=1}^3 (-b x_i - b x_i) + \sum_{i=1}^3 (-a y_1 - a y_2) + \sum_{i=1}^3 (ab + ab) \quad (\text{by the additive rule}) \\ &= \sum_{i=1}^3 x_i (y_1 + y_2) + \sum_{i=1}^3 (-2b) x_i + \sum_{i=1}^3 (-a) (y_1 + y_2) + \sum_{i=1}^3 2ab \\ &= (y_1 + y_2) \sum_{i=1}^3 x_i - 2b \sum_{i=1}^3 x_i - 3a (y_1 + y_2) + (3)2ab \quad (\text{by the multiplicative and constant rules}) \\ &= \left[\sum_{j=1}^2 y_j \right] \sum_{i=1}^3 x_i - 2b \sum_{i=1}^3 x_i - 3a \sum_{j=1}^2 y_j + (3)2ab \quad (\text{by the definition of summation notation}) \\ &= \left[\sum_{i=1}^n x_i \right] \left[\sum_{j=1}^m y_j \right] - mb \sum_{i=1}^n x_i - na \sum_{j=1}^m y_j + (mn)ab \end{aligned}$$

when $n = 3$ and $m = 2$.

Now, let's tackle the fully general case:

$$\begin{aligned} & \sum_{i=1}^n \sum_{j=1}^m (x_i - a)(y_j - b) \\ &= \sum_{i=1}^n \sum_{j=1}^m (x_i y_j - b x_i - a y_j + ab) \\ &= \sum_{i=1}^n \sum_{j=1}^m x_i y_j + \sum_{i=1}^n \sum_{j=1}^m (-b x_i) + \sum_{i=1}^n \sum_{j=1}^m (-a y_j) + \sum_{i=1}^n \sum_{j=1}^m ab \quad (\text{by the additive rule}) \\ &= \sum_{i=1}^n x_i \sum_{j=1}^m y_j - b \sum_{i=1}^n \sum_{j=1}^m x_i - a \sum_{i=1}^n \sum_{j=1}^m y_j + \sum_{i=1}^n \sum_{j=1}^m ab \quad (\text{by the multiplicative rule}) \\ &= \left[\sum_{i=1}^n x_i \right] \left[\sum_{j=1}^m y_j \right] - mb \sum_{i=1}^n x_i - na \sum_{j=1}^m y_j + (mn)ab \quad (\text{by the constant rule}). \end{aligned}$$