R. Nazim Khan
Department of Mathematics and Statistics
nazim.khan@uwa.edu.au

The University of Western Australia

1. Means and Variances

1.1 Summation Notation

Consider data $x_1, x_2, \ldots, x_n$. We use a shorthand notation to write sums as follows:

$$x_1 + x_2 + \ldots + x_n = \sum_{i=1}^{n} x_i.$$

- This is read as "sum from $i = 1$ to n of x_i."
- i is called the *summation index*, and controls what is summed.
- Σ is known as the *summation* notation.

This is also called *sigma notation* as it uses the Greek capital letter sigma.

Note

The summation index i is a place holder, or “dummy variable”, so

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \ldots + x_n = \sum_{j=1}^{n} x_j .$$

Example 1.1

For the data set $x_1 = 1, x_2 = 3, x_3 = -1, x_4 = 4, x_5 = 0, x_6 = -2,$

(a) $\sum_{i=1}^{6} x_i = x_1 + x_2 + x_3 + x_4 + x_5 + x_6 =$

(b) $\sum_{i=1}^{4} x_i =$

Example 1.1 (ctd)

(c) $\sum_{i=1}^{6} x_i^2 = x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 + x_6^2 =$

(d)

$$\sum_{i=1}^{6}(x_i + 4) =$$

$$=$$

Add 4 each time.

(e) $\sum_{i=1}^{6} x_i + 4 =$

Add 4 only once after computing the sum.

Summation Rules

$$\text{S1} \quad \sum_{i=1}^{n} c = nc$$

$$\text{S2} \quad \sum_{i=1}^{n} cx_i = c\sum_{i=1}^{n} x_i$$

$$\text{S3} \quad \sum_{i=1}^{n}(x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$$

$$\text{S4} \quad \sum_{i=1}^{n}(x_i - y_i) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i$$

Proofs

$$\text{S1} \quad \sum_{i=1}^{n} c = \underbrace{c + c + \ldots + c}_{\text{n terms}} = nc$$

$$\text{S2} \quad \sum_{i=1}^{n} cx_i = cx_1 + cx_2 + \ldots + cx_n = c(x_1 + x_2 + \ldots + x_n) = c \sum_{i=1}^{n} x_i$$

$$\begin{aligned} \text{S3} \quad \sum_{i=1}^{n} (x_i + y_i) &= (x_1 + y_1) + (x_2 + y_2) + \ldots + (x_n + y_n) \\ &= (x_1 + x_2 + \ldots + x_n) + (y_1 + y_2 + \ldots + y_n) \\ &= \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i \end{aligned}$$

Proofs (ctd)

S4
$$\sum_{i=1}^{n}(x_i - y_i) = (x_1 - y_1) + (x_2 - y_2) + \ldots + (x_n - y_n)$$
$$= (x_1 + x_2 + \ldots + x_n) - (y_1 + y_2 + \ldots + y_n)$$
$$= \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i$$

Example 1.2

Suppose

$$\sum_{i=1}^{10} x_i = 45, \sum_{i=1}^{10} y_i = 30.$$

Compute the following.

(a) $$\sum_{i=1}^{10}(x_i + y_i) = \sum_{i=1}^{10} x_i + \sum_{i=1}^{10} y_i =$$

(b) $$\sum_{i=1}^{10}(3x_i - 4y_i) = 3\sum_{i=1}^{10} x_i - 4\sum_{i=1}^{10} y_i =$$

Example 1.2 (ctd)

(c)

$$\frac{\sum_{i=1}^{10}(2x_i + 4)}{\left(\sum_{i=1}^{10}(y_i - 2)\right)^2} = \frac{\sum_{i=1}^{10} 2x_i + \sum_{i=1}^{10} 4}{\left(\sum_{i=1}^{10} y_i - \sum_{i=1}^{10} 2\right)^2} \quad (S3, S4)$$

$$=$$

Notes

1. $$\sum_{i=1}^{n} x_i y_i = x_1 y_1 + x_2 y_2 + \ldots + x_n y_n \neq \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)$$
2. $$\sum_{i=3}^{n} c = (n-2)c$$

Note that the last sum starts from 3, so the first two terms are missing, that is, there are total of $(n-2)$ terms to add up.

1.2 Sample Mean

Let $x_1, x_2, \ldots, x_n$ be data from a sample. The *sample mean* is given by

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

Note that this definition implies that

$$\sum_{i=1}^{n} x_i = n\overline{x}, \text{ so } \sum_{i=1}^{n} x_i - n\overline{x} = 0.$$

Some useful results

1. $\displaystyle\sum_{i=1}^{n} (x_i - \overline{x}) = 0$
2. $\displaystyle\sum_{i=1}^{n} (x_i - \overline{x})^2 = \sum_{i=1}^{n} x_i^2 - n\overline{x}^2$

Proof

1. Note that $\overline{x}$ is a constant is these sums, as it does not depend on the summation index i.

$$\begin{aligned}\sum_{i=1}^{n}(x_i-\overline{x}) &= \sum_{i=1}^{n}x_i-\sum_{i=1}^{n}\overline{x}\\ &= \sum_{i=1}^{n}x_i-n\overline{x}\\ &= 0.\end{aligned}$$

Note that to obtain the last line we have used the result

$$\sum_{i=1}^{n}x_i-n\overline{x}=0$$

from the previous slide.

Proof (ctd)

2. We need to multiply out the square term by term. Thus

$$\begin{aligned}
\sum_{i=1}^{n} (x_i - \overline{x})^2 &= \sum_{i=1}^{n} (x_i - \overline{x})(x_i - \overline{x}) \\
&= \sum_{i=1}^{n} \left[x_i (x_i - \overline{x}) - \overline{x}(x_i - \overline{x})\right] \\
&= \sum_{i=1}^{n} \left(x_i^2 - \overline{x}x_i\right) - \overline{x} \underbrace{\sum_{i=1}^{n} (x_i - \overline{x})}_{=0} \\
&= \sum_{i=1}^{n} x_i^2 - \overline{x} \underbrace{\sum_{i=1}^{n} x_i}_{=n\overline{x}} = \sum_{i=1}^{n} x_i^2 - n\overline{x}^2
\end{aligned}$$

Further results

1. $$\sum_{i=1}^{n}(x_i - c) = 0 \Rightarrow c = \overline{x}.$$

2. $\sum_{i=1}^{n}(x_i - c)^2$ is a minimum when $c = \overline{x}$.

Proof

Proof (ctd)

1.3 Sample variance and sample standard deviation

Let $x_1, x_2, \ldots, x_n$ be data from a sample. The *sample variance* is given by

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2\right) \qquad (1)$$

Note that we **always** divide by $n-1$ in the formula for variance.

The positive square root of the sample variance is called the *standard deviation*:

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2} \qquad (2)$$

Computing variance

It is easier to compute variance from data using the formula

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right) \tag{3}$$

Example 1.3 Find the mean and variance of a data set which has summary statistics

$$\sum_{i=1}^{10} x_i = 10, \quad \sum_{i=1}^{10} x_i^2 = 100.$$

Solution