



Summarizing Data: One Variable

The Five Number Summary

In statistics, we're confronted with lots of data, and we need to try to make sense of it all. One way is to summarise the data using the **five number summary**. The five number summary consists of the minimum, first quartile, median, third quartile and maximum of a dataset.

The five number summary:

1. The **minimum** is the smallest observation.
2. The **maximum** is the largest observation in the dataset.
3. When the data are in increasing order, the **median** is the observation halfway from smallest to largest observation.
4. When the data are in increasing order, the **first quartile** is the point that is a quarter of the way.
5. When the data are in increasing order, the **third quartile** is the point three quarters of the way along the sorted data.

A graphical representation of the five number summary is a **boxplot**. To construct a boxplot draw a box extending from the first quartile to the third quartile and indicate the median with a line. Next, attach extensions (or "whiskers") connecting the first quartile to the minimum and the third quartile to the maximum. An example of a boxplot is shown below.

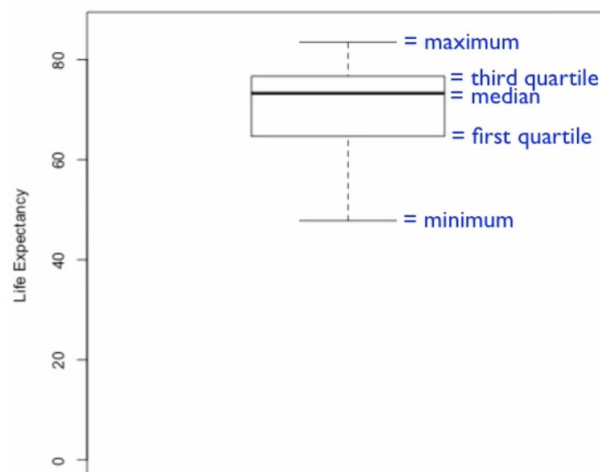


Figure 1: Boxplot with Five Number Summary

EXAMPLE 1

Let's find the five number summary for the following set of 9 numbers:

79, 68, 88, 69, 90, 74, 87, 93, 76

First, rearrange the data in increasing order:

68, 69, 74, 76, 79, 87, 88, 90, 93

The minimum is 68 and the maximum is 93. The median is the number halfway between the minimum and the maximum. Since there are 9 numbers the median is the number at the fifth position corresponding to 79. The first quartile is the number one quarter of the way from the first number up to the ninth number. In this case this is the third number, equal to 74. For the third quartile, we take the number at three quarters along from the smallest number to the largest, equal to 88.

EXAMPLE 2

Let's try one more example of computing the five number summary. Suppose the data are the following 6 numbers:

78, 93, 68, 84, 90, 74

Again, we start by arranging the data from the smallest up to the largest:

68, 74, 78, 84, 90, 93

We see that the minimum value is 68, and the maximum value is 93. But now what about the median? That should be the number which is halfway from the first number up to the sixth number. In this case, that's the number at position 3.5. There is no such number so instead we calculate what number is halfway between the third (78) and fourth (84) number.

$$\frac{84 - 78}{2} = \frac{6}{2} = 3 \qquad \text{median} = 78 + 3 = 84 - 3 = 81$$

Now, what about the quartiles? The first quartile should be the number which is one-quarter of the way along from the first number up to the sixth number. To calculate the position of quartiles in the sorted data, we are using the formula

$$1 + p(n - 1)$$

where n is the number of data values and p is the desired percentile. For the first quartile, $p = 0.25$ and for this example, $n = 6$, so the first quartile is the data value at the *2.25th* position in the sorted data.¹ Again, this is not an integer position so we must calculate

¹There are multiple ways to interpolate between the data values in order to estimate the quartiles. All of the methods give similar values. We're using a common method which has nice theoretical properties (related to the "cumulative distribution function", which we won't be covering in this course).

the number a quarter of the way between the second number (74) and the third number (78).

$$(78 - 74) \times 0.25 = 1$$

$$\text{first quartile} = 74 + 1 = 75$$

Similarly, for the third quartile, we need the number which is three-quarters of the way from the first number to the sixth number, which in this case is the number at the *4.75th* position. This corresponds to the number three-quarters of the way between the fourth (84) and the fifth numbers (90).

$$(90 - 84) \times 0.75 = 4.5$$

$$\text{third quartile} = 84 + 4.5 = 88.5$$

No matter how many numbers are in a dataset, we can compute the five number summary.