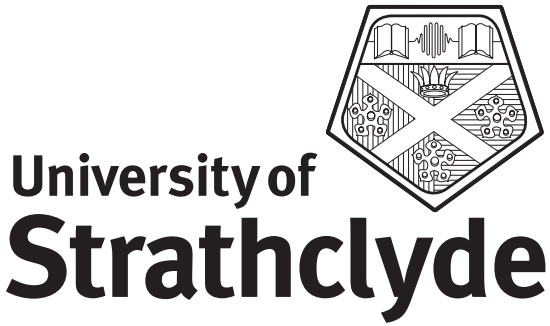

MM104/ MM106/ BM110

Topic 1: Data types, Central Location and Missing Values
Missing Values

Ainsley Miller
ainsley.miller@strath.ac.uk

Missing Values

In statistics, missing data, or missing values, occur when no data value is stored for the variable in an observation. Missing data are a common occurrence and can have a significant effect on the conclusions that can be drawn from the data.
In this section we will show you how to calculate missing values of the mean and median.

Missing Values Involving the Mean - Example

The best way to illustrate how find a missing value is through an example.

The mean score on a calculus test was 70 %. The lecturer can only find 8 scripts, but 9 students sat the test. The scores of the 8 students were: 71, 39, 85, 97, 63, 46, 80 and 69. What is the score of the missing student ?

The formula for the mean from the previous lecture is:

$$\bar{x} = \frac{x_1 + x_2 + \ldots, x_n}{n}$$

From the question $\bar{x} = 70, n = 9$ and $x_1, x_2, \ldots x_n$ are

$71, 39, 85, 97, 63, 46, 80, 69, x_9$

Missing Values Involving the Mean - Example

Substituting in obtains:

$$\Rightarrow 70 = \frac{71 + 39 + 85 + 97 + 63 + 46 + 80 + 69 + x_9}{9}$$

$$\Rightarrow 630 = 71 + 39 + 85 + 97 + 63 + 46 + 80 + 69 + x_9$$

$$\Rightarrow 630 = 550 + x_9$$

$$\Rightarrow x_9 = 80$$

The grade of the missing student is 80 %

Missing Values Involving the Median

Finding the missing value when you are given a median is a bit more complicated.

The approach changes dependent on whether or not you have an odd or even sample size.

We will firstly look at an odd sample size as this is the more straightforward.

Missing Values Involving the Median - Odd Sample Size

The median from a data set is known to be 13. The data set is comprised of the following values: 7, 2, 15, 20, 8 and 14. What is the missing value ?

The total sample size is 7 (including the missing value)

Arrange the data in ascending order

$2, 7, 8, 14, 15$ and 20

Since the sample size is an odd number, the median is the middle number, and the only way to find the missing number is if the median is the missing number.

The median is 13

I think I need more convincing

Do you need more convincing ?

Well let's check our answer.

Given we think the missing value is 13. Let's put that into our data set and find the median and confirm it is in fact 13.

Firstly put the data in ascending order: $2, 7, 8, 13, 14, 15$ and 20

The median lies in the $\dfrac{n+1}{2}$, position in the data. Here $n = 7$

$$\Rightarrow \text{Position } = \frac{7+1}{2} = \frac{8}{2} = 4$$

The 4th value in the data set is the median which is ... 13.

Missing Values Involving the Median - Even Sample Size

The median from a data set is known to be 8. The data set is comprised of the following values: 1, 3, 4, 10, 11, 17 and 20. What is the missing value ?

The data is already in ascending order

The median lies in the $\frac{n+1}{2}$, position in the data. Here $n = 8$

$$\Rightarrow \text{Position} = \frac{8+1}{2} = \frac{9}{2} = 4.5$$

The median lies in between the 4th and 5th value.

Missing Values Involving the Median - Even Sample Size

The 4th element is 10

The 5th element is 11

The median is: $10 + \dfrac{11 - 10}{2} = 10 + \dfrac{1}{2} = 10.5$

But we were told from the question that the median is 8

what has went wrong ?

We have not taken into account the missing value !

The calculated median is too high,

The missing value must be located in between the 3rd and 4th element

Let's denote the missing value by x

We have: 1, 3, 4, x, 10, 11, 17 and 20

We know the median is 8 and is located in between the 4th and 5th element

$$\text{median} = \text{lower value} + \frac{(\text{upper value - lower value})}{2}$$

$$\Rightarrow 8 = x + \frac{10 - x}{2} \qquad \text{Multiply by 2 on both sides}$$

$$\Rightarrow 16 = 2x + 10 - x$$

$$\Rightarrow 16 = 10 + x$$

The missing value is 6.