

Calculating Some Summary Statistics with Excel

I. Mean and Standard Deviation:

Example 1:

What is the average waiting time for a bus? Randomly sample 12 different days. The data are:

10 minutes
6 minutes
8 minutes
23 minutes
6 minutes
14 minutes
17 minutes
10 minutes
26 minutes
20 minutes
5 minutes
5 minutes

The sample mean is:

$$(10 + 6 + 8 + 23 + 6 + 14 + 17 + 10 + 26 + 20 + 5 + 5) / 12 \\ = 150 / 12 = 12.5 \text{ minutes}$$

Suppose you want to calculate the mean, and you entered the data in an Excel spreadsheet as follows:

10 in cell C1
6 in cell C2
8 in cell C3
23 in cell C4
6 in cell C5
14 in cell C6
17 in cell C7
10 in cell C8
26 in cell C9
20 in cell C10

5 in cell C11
5 in cell C12

To calculate the average waiting time, go to the Function Wizard. The function you are going to use is in the category of STATISTICAL and is the function AVERAGE. You will have to enter the data range (the data is in C1:C12)

NUMBER 1: C1:C12

For the above problem, the average wait is 12.5 minutes.

Note: To calculate the standard deviation, go to the Function Wizard. The function you are going to use is in the function category of STATISTICAL and is STDEV.S. This estimates the standard deviation based on a sample. [If you take a census, then you use STDEV.P. The difference between the two formulas is whether you divide by N (the size of the population) or n-1 (the sample size minus one).] Most of the time, researchers are working with samples so the STDEV.S function is used.

Again, you have to enter the data range.

NUMBER 1: C1:C12

The answer I got for the above data set is 7.39 minutes.

II. MEDIAN

To get the median by hand, you first have to order the data from lowest to highest. The median is the number exactly in the middle. If the sample size, n , is even, then the median is the average of the two middle numbers.

Definition: The median is that number for which half the data is larger than it, and half the data is smaller. It is also called the 50th percentile.

Using the bus waiting times dataset, let's put them in order:

5
5
6
6
8
10
10
14
17
20
23
26

Since there is an even number of observations (12), the median is exactly in between the 6th and 7th ordered numbers, or, the average of 10 and 10. 😊

The median is 10.0 minutes. Half of the time, the sample wait was more than 10 minutes, and half of the time it was less.

In Excel, you do not have to order the data yourself. To calculate the median waiting time, go to the Function Wizard. The function you are going

to use is in the category of STATISTICAL and is the function MEDIAN. You will have to enter the data range (C1:C12):

NUMBER 1: C1:C12

The median is 10.0 minutes.

Example 2:

A company is studying absenteeism among its employees in one particular plant. They collect the following data:

Absences over a one month period for 15 employees:

6 absences

5 absences

1

1

1

3

4

2

2

2

0

0

1

1

1

We will calculate the mean and the median for this dataset, first by hand, and then using Excel.

Mean = $(6 + 5 + \dots + 1) / 15 = 30 / 15 = 2.0$ days

The average employee in this group was absent 2 days during this one-month period.

To calculate the median, first order the data:

0
0
1
1
1
1
1
1
1
2
2
2
3
4
5
6

Since the number of observations, $n=15$, is odd, the median is exactly at the center, at the 8th ordered observation. The median is 1 day.

In Excel,

Enter the 15 data items into a spreadsheet at, say, D3 through D17. Using the function wizard, choose AVERAGE from the STATISTICAL category and enter the range, D3:D17. The average is 2.0. Make sure to format the cell where you place the mean, so that it can take decimal places. Most problems need 2 decimal places.

Median: Use the function wizard to choose MEDIAN from the STATISTICAL category and enter the range D3:D17. The median is 1.0 day.