C3 … Standard Deviation

Standard Deviation is a measure of spread in Statistics. It is used to quantify the measure of spread OR variation of a set of data values.

A low measure of Standard Deviation indicates that the data are less spread out, whereas a high value of Standard Deviation shows that the data in a set are spread apart from their mean average values.

Standard Deviation is calculated by :

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

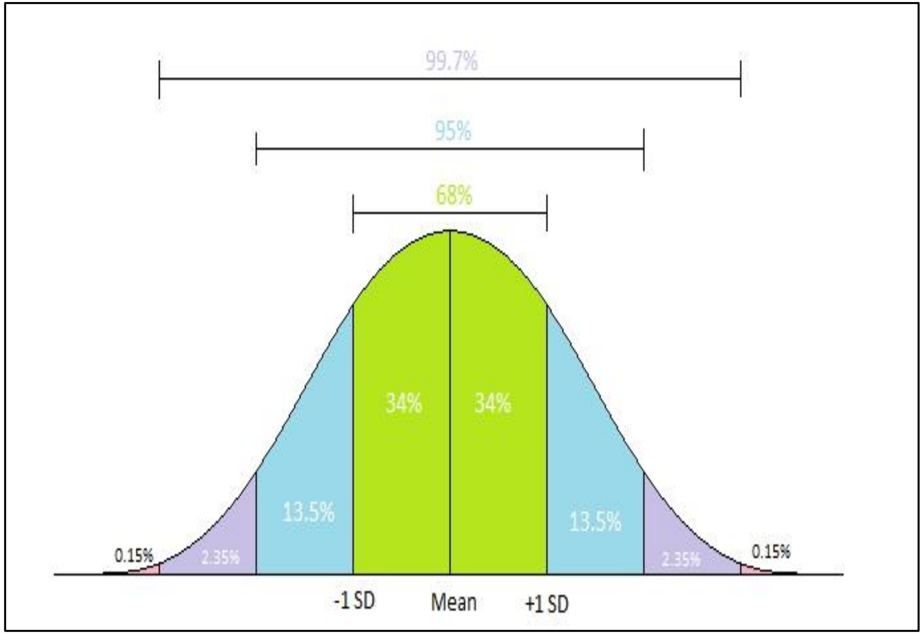
where x1, x2, x3.....xn are observed values in sample data,
$\overline{x}$ is the mean value of observations and
N is the number of sample observations.

Mean = 5.7

Sum all Deviation_sq
And then divide by (N-1)
and then square_root

List	deviation_sq
9	$(9-5.7)^2$
10	
1	
2	
3	
4	
5	
6	
7	
8	
8	

Confidence Interval …(1)



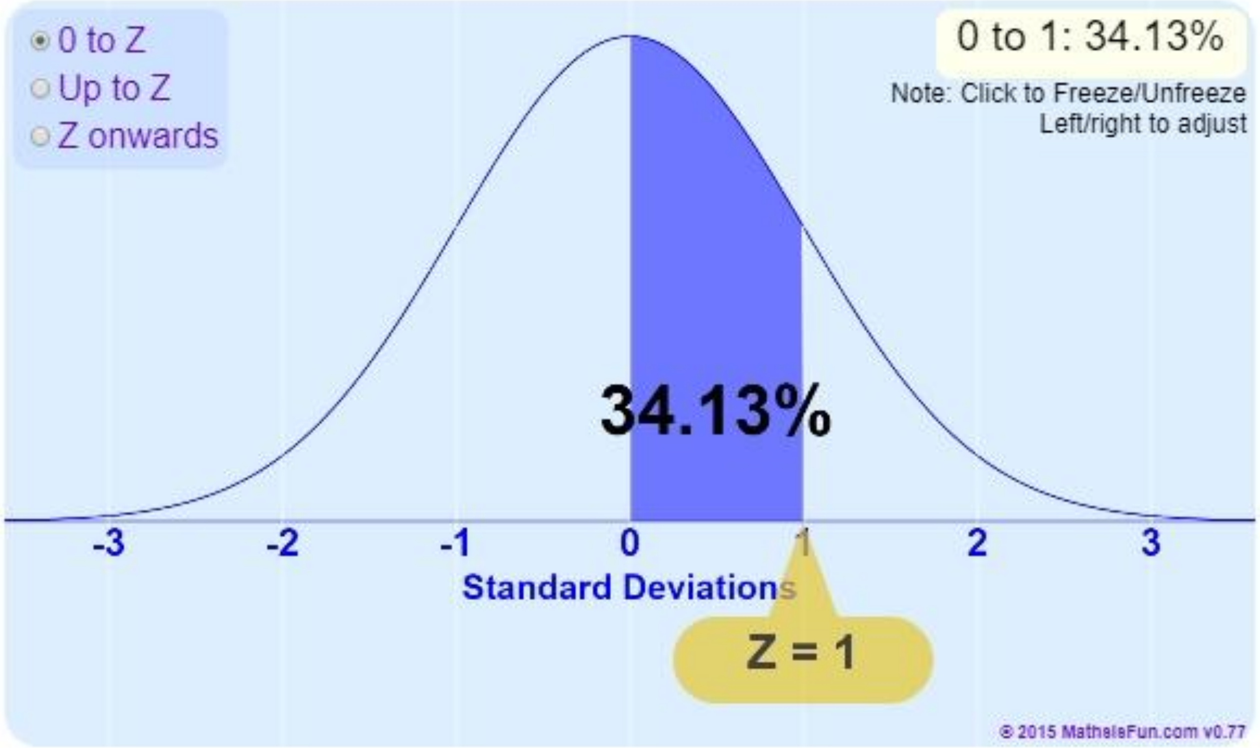
The Table

You can also use the table below. The table shows the area from 0 to Z.

Instead of one LONG table, we have put the "**0.1**"s running down, then the "**0.01**"s running along. (Example of how to use is below)

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

Confidence Interval …(2)



This is the "bell-shaped" curve of the Standard Normal Distribution.
It is a Normal Distribution with mean 0 and standard deviation 1.

Confidence Interval …(3)

Calculating the Confidence Interval

Step 1: find the number of observations **n**, calculate their mean $\bar{X}$, and standard deviation **s**

Using our example:

- Number of observations: **n = 40**
- Mean: $\bar{X}$ **= 175**
- Standard Deviation: **s = 20**

> Note: we should use the standard deviation of the entire **population**, but in many cases we won't know it.
>
> We can use the standard deviation for the **sample** if we have enough observations (at least n=30, hopefully more).

Step 2: decide what Confidence Interval we want: 95% or 99% are common choices. Then find the "Z" value for that Confidence Interval here:

Confidence Interval	Z
80%	1.282
85%	1.440
90%	1.645
95%	1.960
99%	2.576
99.5%	2.807
99.9%	3.291

For 95% the Z value is **1.960**

Step 3: use that Z in this formula for the Confidence Interval

$$\bar{X} \pm Z\frac{s}{\sqrt{n}}$$

Confidence Interval …(4)

Another Example

Example: Apple Orchard

Are the apples big enough?

There are hundreds of apples on the trees, so you randomly choose just **46** apples and get:

- a Mean of **86**
- a Standard Deviation of **6.2**

So let's calculate:

$$\overline{X} \pm Z\frac{s}{\sqrt{n}}$$

We know:

- $\overline{X}$ is the mean = 86
- **Z** is the Z-value = 1.960 (from the table above for 95%)
- **s** is the standard deviation = 6.2
- **n** is the number of observations = 46

$$86 \pm 1.960 \times \frac{6.2}{\sqrt{46}} = \mathbf{86 \pm 1.79}$$

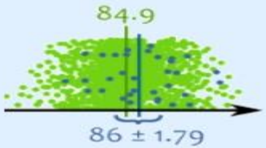
So the true mean (of all the hundreds of apples) is **likely** to be between 84.21 and 87.79

True Mean

Now imagine we get to pick ALL the apples straight away, and get them ALL measured by the packing machine (this is a luxury not normally found in statistics!)

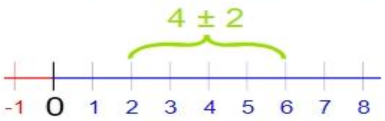
And the **true mean** turns out to be **84.9**

Let's lay all the apples on the ground from smallest to largest:



Confidence Interval …(5)

Confidence Intervals



An interval of **4 plus or minus 2**

A Confidence Interval is a **range of values** we are fairly sure our **true value** lies in.

Example: Average Height

We measure the heights of **40** randomly chosen men, and get a mean height of **175cm**,

We also know the standard deviation of men's heights is **20cm**.

The **95% Confidence Interval** (we show how to calculate it later) is:

175cm ± 6.2cm

This says the **true mean** of ALL men (if we could measure all their heights) is likely to be between 168.8cm and 181.2cm.

But it might not be!

The "95%" says that 95% of experiments like we just did will include the true mean, but **5% won't**.

So there is a 1-in-20 chance (5%) that our Confidence Interval does NOT include the true mean.