**Z-SCORE BASICS OF CALCULATING AND ANALYSING**



In statistics one most important concept is **Z-score** widely used in **DESCRIPTIVE STATISTICS AND INFERENTIAL STATISTICS.**



We often, got confused when to use Z-SCORE and what formula to be used!

Today all your doubts will be cleared regarding all above mentioned dilemma.

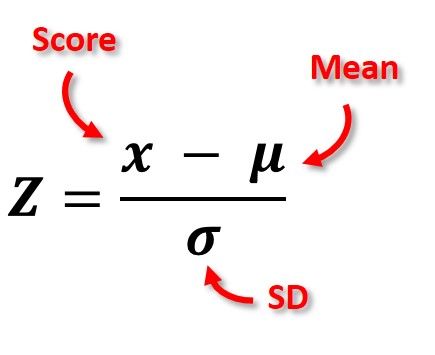
WHAT IS Z-SCORE?

**A z-score describes the position of a raw score in terms of its distance from the mean, when measured in standard deviation units. The z-score is positive if the value lies above the mean, and negative if it lies below the mean.**

**It is also known as a standard score, because it allows comparison of scores on different kinds of variables by standardizing the distribution.**

FORMULA INTERPRETATIONS

Firstly, start with the formula generally used in calculating z-score.



The above formula has two interpretations (only for understanding).

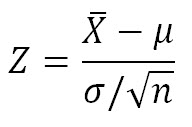
1. As population standard deviation (σ) is given, so we can assume we are calculating z for population value. (Usually for descriptive statistics)
2. Generally, the formula of Z-score with standard error (used mostly in inferential statistics)

We can derive the same above formula by just putting n = 1 which means that we are calculating Z-score for each and every individual data values.

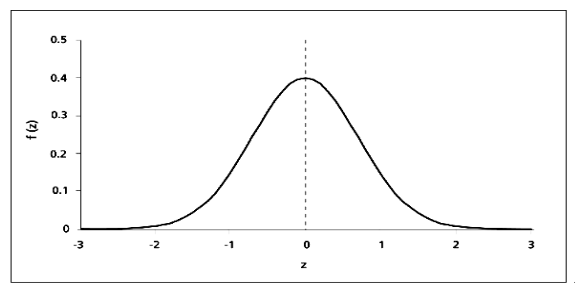
Formula given below:

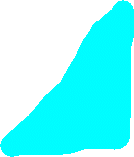
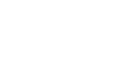
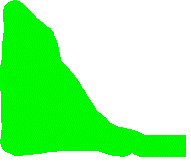
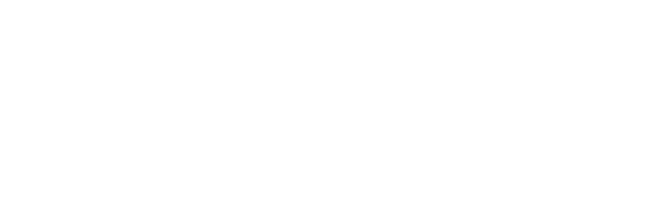
This implies that sample from population have some error.





A **standard normal distribution (SND)** is a normally shaped distribution with a **mean of 0 and a standard deviation (SD) of 1 see below figure.**





**blue: left or negative from mean**

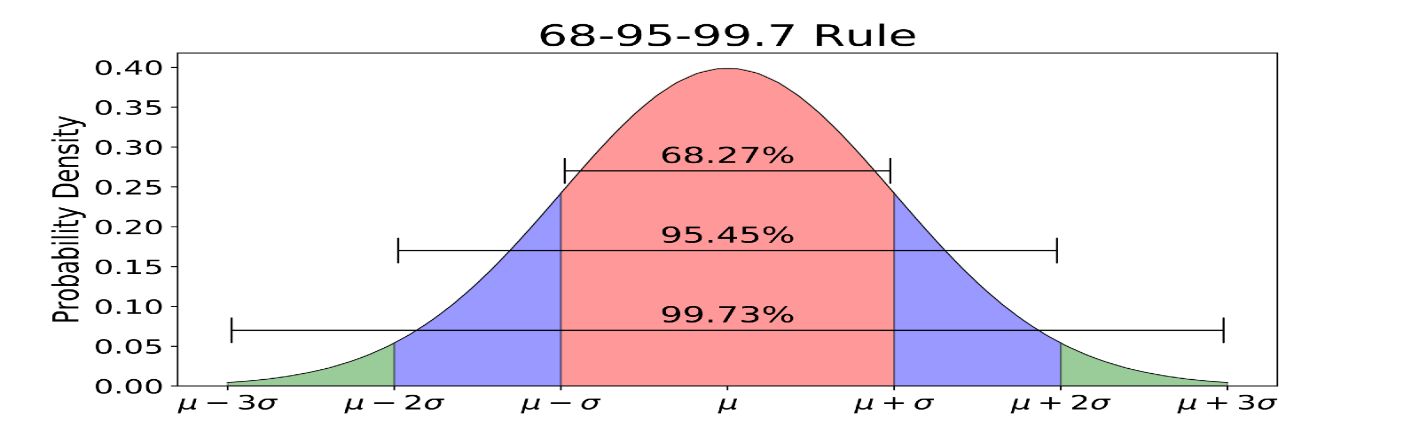
**green: right or positive from mean value**

How do you interpret a z-score?

The value of the z-score tells you how many standard deviations you are away from the mean. If a z-score is equal to 0, it is on the mean.

A **positive z-score** indicates the raw score is **higher than the mean average**. For example, if a z-score is equal to +1, it is 1 standard deviation above the mean.

A **negative z-score** reveals the raw score is **below the mean average**. For example, if a z-score is equal to -2, it is 2 standard deviations below the mean.

****



**empirical rule**

|  |
| --- |
| Green one is more spread way from mean that is., 3 standard deviation away from mean. Area is 99.73%  Purple is less spread than green one but still away from mean that is., standard deviation 2 away from mean. Area is 95.45%  Pink is close and is less spread from mean as compared to all purple and green shaded regions. Area is 68.27% |

**Above is an empirical rule of Normal or Gaussian Distribution.**

It is a standard one Indicating more clearly how much standard deviation is deviating and what area under it falls explained in box above.

In other words, The Standard Normal Distribution **shown by fig-1 below and above empirical rule** *allows researchers to calculate the probability of randomly obtaining a score from the distribution (that is, sample). For example, there is a 68% probability of randomly selecting a score between -1 and +1 standard deviations from the mean.*

*The probability of randomly selecting a score between -1.96 and +1.96 standard deviations from the mean is 95%. If there is less than a 5% chance of a raw score being selected randomly, then this is a statistically significant result.*

Here we use z-score to know about deviations in descriptive statistics and for testing whether to accept the model or not by accepting and rejecting null hypothesis according to condition (in inferential statistics).

We knew about 1, 2 and 3 standard deviation away from mean area but what about 4,5, 6 ,….,n deviations from mean area.

We calculate Z-score

How to calculate z-score?

Before calculating, let us understand about the structure how to see values from Z-table and its interpretation.

The **table given** above is designed specifically for standard normal distribution. The mean of these tables is 0 and 1 is their standard deviation.

In the **first column of the table**, we can find out the number of standard deviations either above or below the mean value to one decimal place. (The integer part and the first decimal of Z-score are present in the row label).

Across the **topmost row of the table**, the part which denotes the z-score denotes the hundredth. Then, the intersecting point of the columns and rows provides us with the area under the normal curve or the probability.

**Let us consider one numerical**

**Imagine a group of 400 applicants who took a math test. zen got 720 points (X) out of 1000. The average score was 600 (µ) and the standard deviation was 150 (σ). Find out how well zen performed compared to his peers**.

Solution:

From the above-given data, we can deduce that:

The value of x = 720

The value of μ = 600 And the value of σ = 150

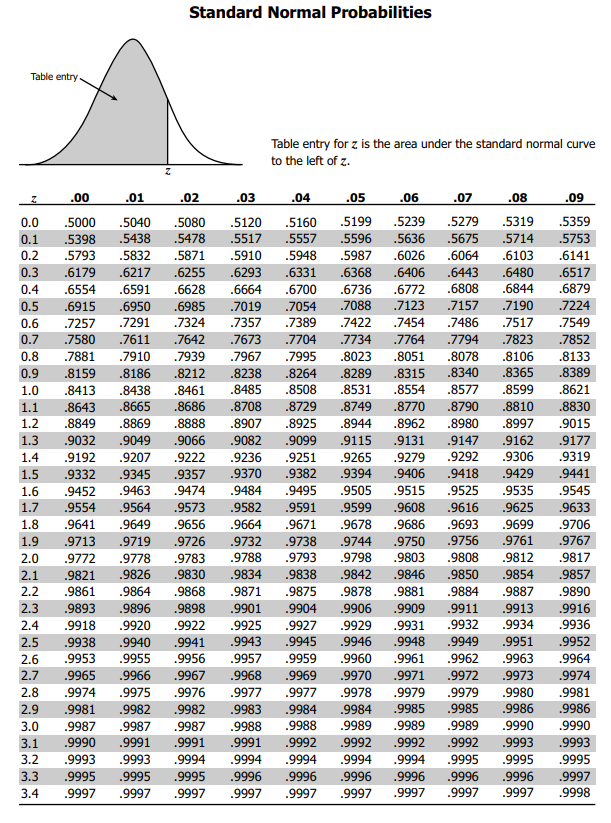
Z score = (x – μ) / σ  = (720-600) / 150   = 0.80

Using the Z-score table we can find out how well she performed relative to her peers. Now we need to determine the percentage of peers whose score goes higher and lower than the scores of Zen.

In this example the Z-score calculated is positive, therefore we refer to all the positive values in the Z-score table. There are certain steps to be followed while using the Z score table.

### **Steps to be Followed While Referring to the Z-scale Table**

1. *First, find the first two digits on the y-axis (in our example the first two digits are 0.8).*
2. *Then, go to the x-axis in order to find the second decimal number (according to our example it is 0.0) the number is .7881*





**z-table**

Now, if question changes we have to find the area or probability under 3 situations

1. Lower than 85 2. Higher than 85 3. Between 85 and 100

Given, average mean as 100 and standard deviation as 15

**Solution:**

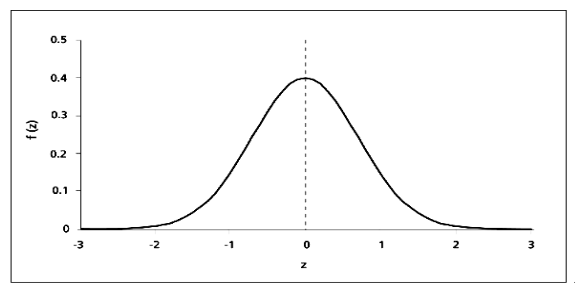
1. Z score = (x – μ) / σ  = 85 - 100/15 = 0.1587 (from z-table) or 15% (approx.)

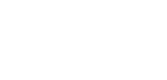
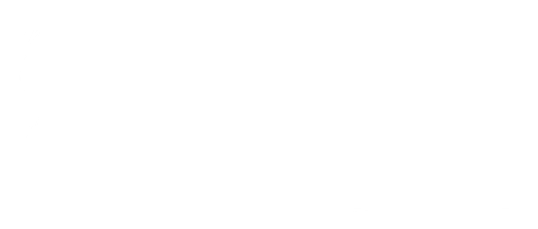
|  |
| --- |
| Note: a. Z-table always provide probability area from left of the respective value.  See the source image  b. Total area under the curve will be one. And it is symmetric dividing curve into two parts equally that is 50% each. |

1. now we know complete area as 1 then we calculate this as

p(x>85) = 1- p(x<= 85) = 1 – 0.1587 (calculated above) = 0.8413 or 84% (approx.)

1. p( 85<x<100) = 0.50 – 0.1587 = 0.3413 or 34%.



**APPLICATION OF Z-SCORE**

In consideration of observed and measured values, the row represented in a line is either above or below the mean values. This is termed a standard score which is a number of standard deviations. These are most commonly called Z scores. The other terms which are used in place of z score are z values, normal scores, standardized variables, and Pull high energy physics.

In order to calculate the z score, one must have proper knowledge of the mean and standard deviation of the complete population.

A number of z-score applications are used widely over a range of fields. This can be enlisted as described below:

1. **Z test**:

Latest is generally used in the standardization of the testing which can be considered as an analogue of student's t-test also. Most of the time it becomes very difficult to calculate the entire population so the t-test is much more widely used as it is more precise.

1. **Comparison of scores measured on different scales that are ACT and SAT**:

In order to observe the performance of the students a very old process is often used in High schools’ tests which are known as SAT and ACT. To use this particular assessment test, it becomes very convenient to compare a student's score. This process also helps to show the main and standard deviation for the total score on the SAT and ACT.

1. Percentage and observation below a z score:

This application is much related to the previous application only. The results of SAT and ACT are again taken into consideration in order to compare the scores of the test takers.

1. The relative importance of variables in multiple regression: standardized regression coefficients