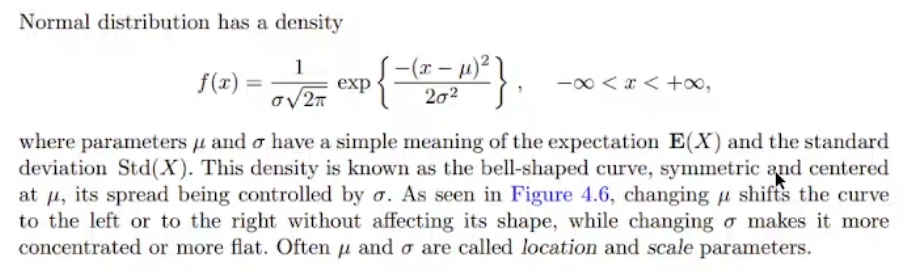
Normal Distribution

Distribution we see a lot in nature.



1/() ---------> coefficient

Graphical user interface, text

Description automatically generated

Chart, histogram

Description automatically generated

Right graph is continuous version of normal distribution. It is defined continuously but in real life, even in some discrete cases, normal distribution can show off.

If you use interpolation and make it continuous, you end up with graph on the right.

We collect discrete data and use interpolation to make it continuous function.

Graphical user interface, application

Description automatically generated

Standard deviation is same in left graph. How values are distributed is still the same in left graph.

Mean is same so top points and boundaries (start and end points) are same in right graph. Meaning of standard deviation is distance between data point and mean. So distribution of values are changed.

Area under the curve doesn’t change in both pictures, always 1.

f(x) never become 0 if you look at the function definition. So any of the curves don’t touch the x axis.

Chart, line chart

Description automatically generated

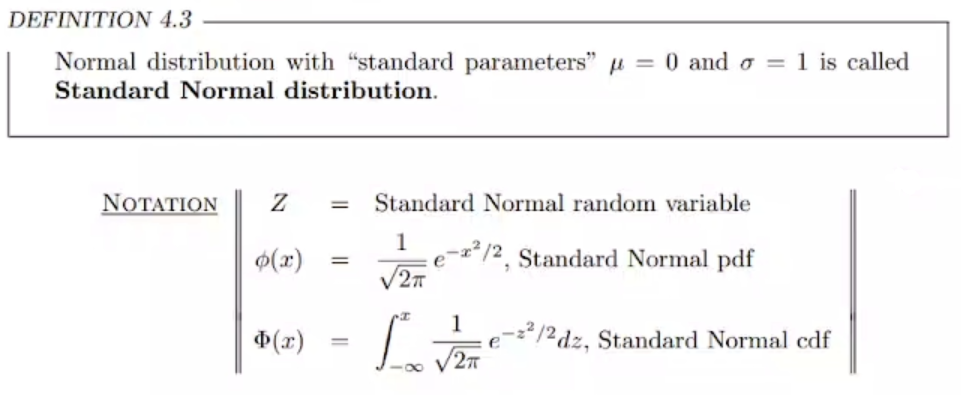
If you know one parameter you can find other one bc area needs to be 1. Parameters are M and SD.

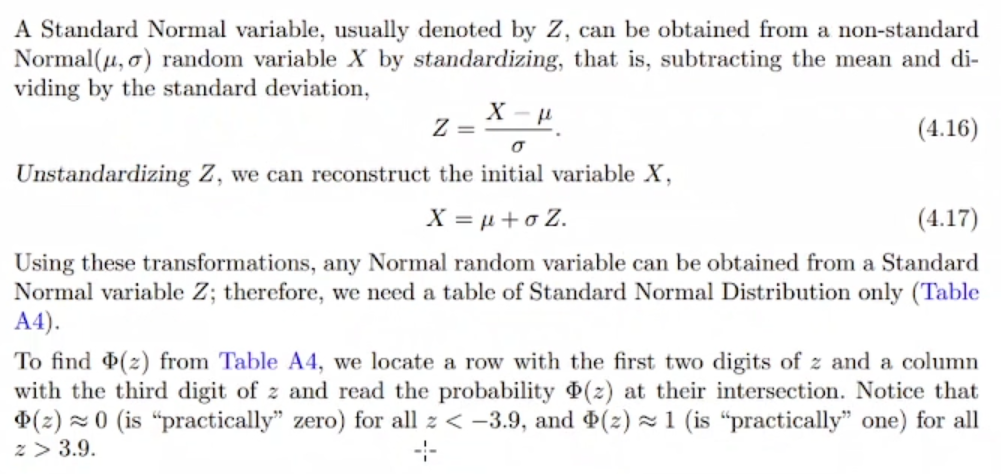
Chart, line chart

Description automatically generated

With table, we can find area up until some point. So you will find the area until 1380 and then subtract it from 1.

Standard Normal Distribution



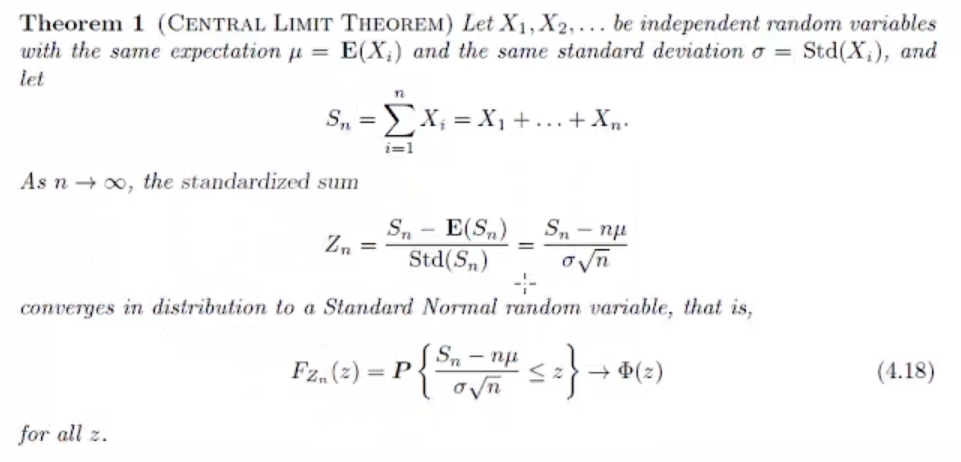


To standardize X, you move it by mean to put it into origin and divide to sd to change how wide of the wings of bell. Then we have standardized the point. Using transformation, any normal distribution variable can be obtained from a Z.



Percents are always like these for standard normal distribution.

Central Limit Theorem

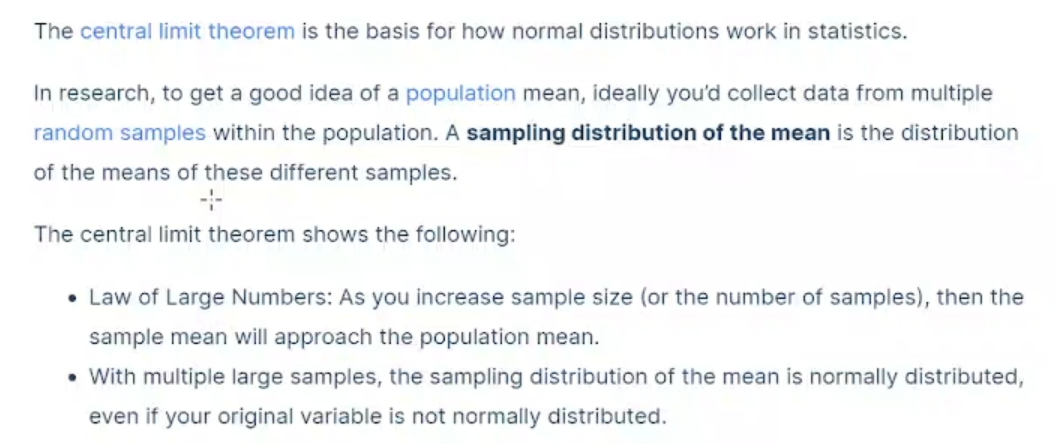


n 🡪 # of random variables.

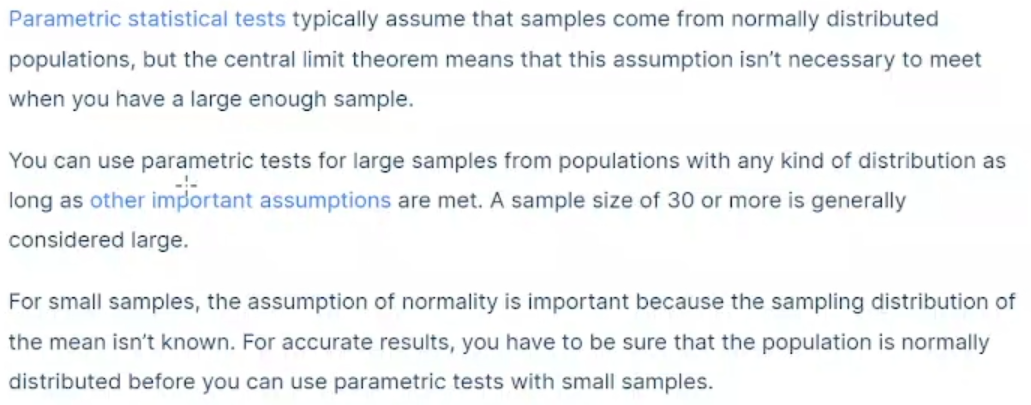
Zn is substituted version of the original transformation.

We just have infinitely many random independent variables, we don’t know distribution of those. All we know their expectations and sd are same. If you use transformation we have used before, we get to standard normal distribution at the infinity.

If you have large number of data with same M and SD, then you should have standard normal distribution.



If you have large enough sample space, then mean of the sample will approach the population mean.



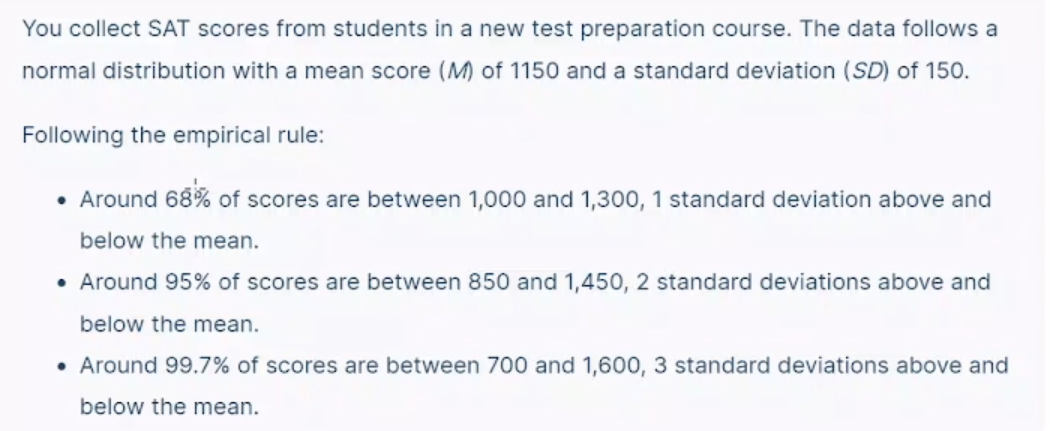
Lets say you do survey and you don’t have enough number of data. In that case, you just assume that this is normally distributed. Your assumption should be based on some logical background.

If you have large enough sample space, you will most likely see that data will be distributed normally and you have to show that.

You want error is minimal and data accumulated around the mean. Then we can convince other people.

Most of the stuff we measure in the nature is normally distributed.

Example:



Chart, histogram

Description automatically generated

You only need to know the mean and standard deviation of your distribution to find the z-score of a value.

A picture containing graphical user interface

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

We use letter z because of z-score or z-value.

When you try to convince your boss to validity of your data, they are gonna ask you “Are they normally distributed?”. If you say yes, they will ask for z-score or p-score (we will mention p later).

z-score means how you use the transformation.

Standard version of the table has a lot better numbers.

If you make sth standard, it means you want to compare that with sth else.

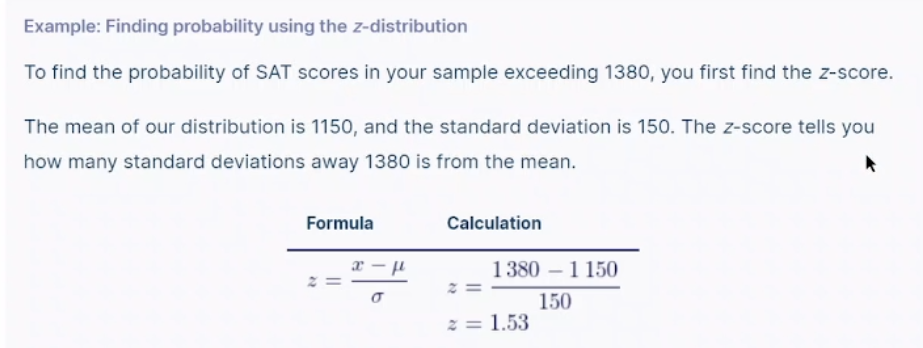
Finding probability using the z-distribution

Text

Description automatically generated

p-value is probability value that is on the curve. You find the area up until some point. That probability value you look at certain point is called p-value if that is associated with z-value.

Example:



x is non-standard value.

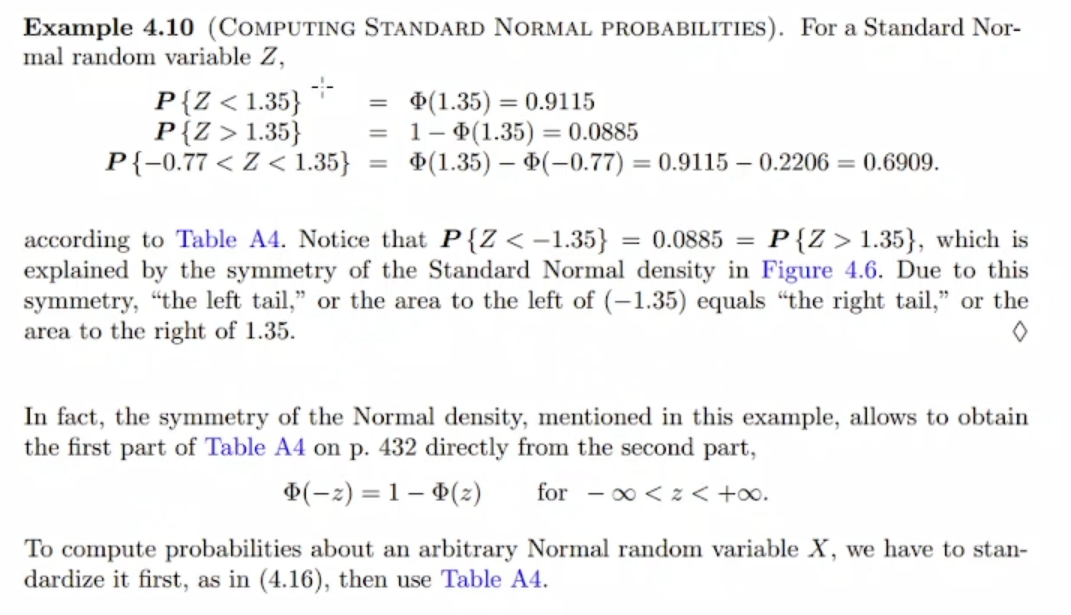
1.53 is valid number. In standard normal distribution values are scattered around 0 bc mean is 0. If value is between 1 and -1, it means it is 68%. So if you find your z-value close to 1, it is good sign for your data.

Chart, line chart, histogram

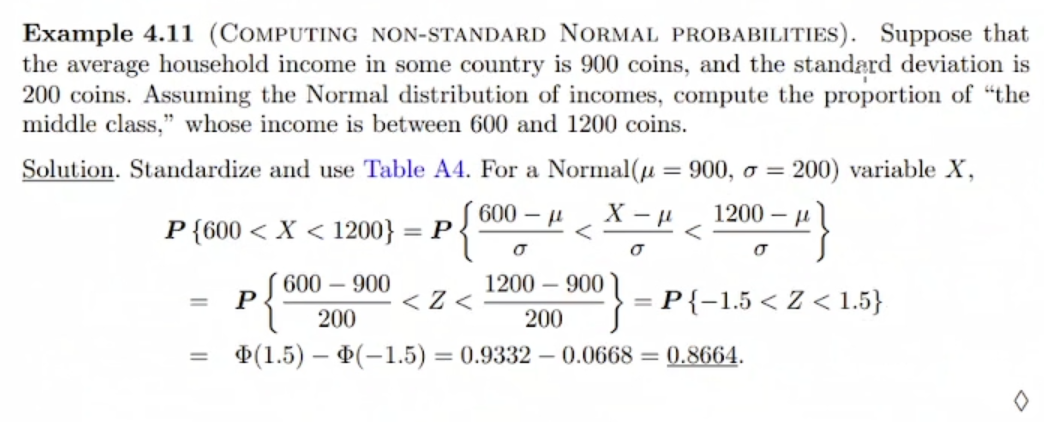
Description automatically generated

p-value is y coordinate of z-value.

Example:



Example:



You can also do it with integral and calculations if you don’t want to use table.

Standard normal distribution is more understandable and easier to calculate.

When you standardize, you understand that you are looking at some symmetric interval (-1.5 to 1.5). So we are actually looking at the middle class.

Example:

