Topic 11 d: Probability: Normal Distribution

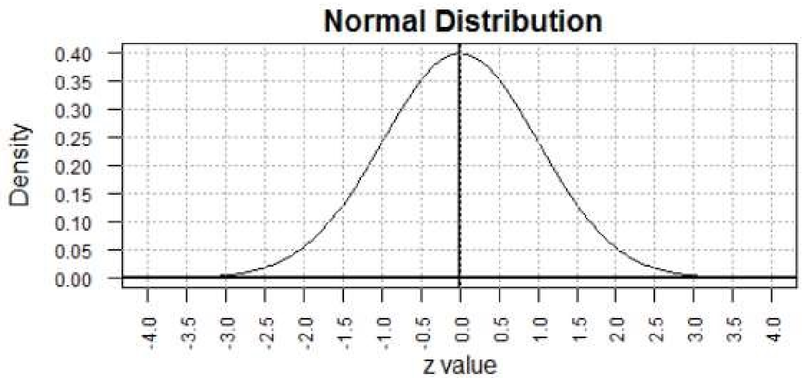
Not that you need to know it but it is nice to see the mathematical definition of the Normal Distribution. First the big picture, for any μ and σ.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Then, the more simple version when μ = 0 and σ = 1

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Then here is the standard graph of a Normal Distribution with μ = 0 and σ = 1

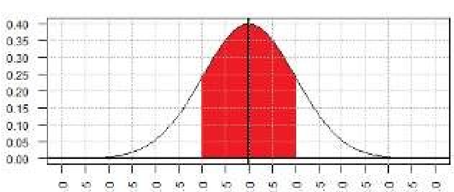


The area under the curve and above the z-axis is **1** square unit. The P(X < 1.3) is the area under the curve, above the z-axis, and to the left of the vertical line at z=1.3.

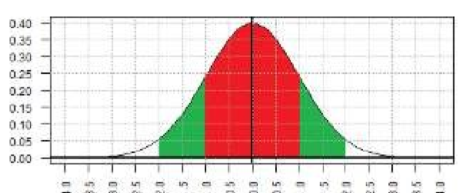
One essential point that is not obvious from the graph is that the graph of the function never gets to actually touch the z-axis. The curve continues to the left and right, forever. However, there is very little area between the curve and the z-axis at those extreme values.

A point that is obvious from the graph is that the Normal Distribution is symmetric. For this graph, because μ=0 and σ=1, the z-values represent a number of standard deviations as we move away from the 0 value.

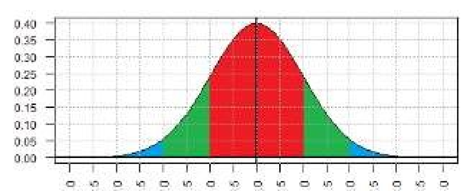
About 68% of the area is between -1 and 1.



About 95% of the area is between -2 and 2.



About 99.7% of the area is between -3 and 3



We do have a nice table that gives the cumulative probability for the Normal Distribution when we have μ=0 and σ=1. See the link on the web page. We could use the table to solve

P(X < 1.45) =

P(X < -0.28) =

P(X > -1.23) =

P(X > 0.66) =

P(X < 1.78 or X > 2.13) =

P(-0.57 < X < 1.03) =

Clearly, it is a pain to use the table. Fortunately, we have the pnorm() function that will solve these same problems.

```
     #  find  P X < 1.45 )
pnorm( 1.45 )
     #  find P( X  < -0.28 )
pnorm( -0.28 )
     #  find P( X > -1.23 )  but that  = 1 - P( X < -1.23)
1 - pnorm( -1.23 )   # old way
pnorm( -1.23, lower.tail = FALSE)    # better way
     # find P( X > 0.66 ) but that is 1 - P(X < 0.66 )
1 - pnorm( 0.66 )   # old way
pnorm( 0.66, lower.tail=FALSE )   # better way
```

```
>      #  find  P X < 1.45 )
> pnorm( 1.45 )
[1] 0.9264707
>      #  find P( X  < -0.28 )
> pnorm( -0.28 )
[1] 0.3897388
>      #  find P( X > -1.23 )  but that  = 1 - P( X < -1.23)
> 1 - pnorm( -1.23 )   # old way
[1] 0.8906514
> pnorm( -1.23, lower.tail = FALSE)    # better way
[1] 0.8906514
>      # find P( X > 0.66 ) but that is 1 - P(X < 0.66 )
> 1 - pnorm( 0.66 )   # old way
[1] 0.2546269
> pnorm( 0.66, lower.tail=FALSE )   # better way
[1] 0.2546269
```

```
     # find P( X < 1.78 or X > 2.13 )
     # that is same as  P(X < 1.78 ) + P( X > 2.13 )
     # that is same as  P(X < 1.78 ) + (1 - P( X< 2.13 ) )
pnorm( 1.78 ) + (1 - pnorm( 2.13 ) )  # old way
pnorm( 1.78 ) + pnorm( 2.13, lower.tail=FALSE)  # better way
     # find P( -0.57 < X < 1.03 )
     #that is find P( X < 1.03 ) - P( X < -0.57 )
pnorm( 1.03 ) - pnorm( -0.57 )
```

```
>      # find P( X < 1.78 or X > 2.13 )
>      # that is same as  P(X < 1.78 ) + P( X > 2.13 )
>      # that is same as  P(X < 1.78 ) + (1 - P( X< 2.13 ) )
> pnorm( 1.78 ) + (1 - pnorm( 2.13 ) )  # old way
[1] 0.9790478
> pnorm( 1.78 ) + pnorm( 2.13, lower.tail=FALSE)  # better way
[1] 0.9790478
>      # find P( -0.57 < X < 1.03 )
>      #that is find P( X < 1.03 ) - P( X < -0.57 )
> pnorm( 1.03 ) - pnorm( -0.57 )
[1] 0.5641561
```

Now that we have looked at those problems, consider problems that require us to read the table "backwards." First, here are such problems and we will solve them by using the table.
Find y such that $P(X < y) = 0.23$

Find y such that $P(X > y) = 0.075$

Find y such that $P(X < -y \text{ or } X > y) = 0.156$

Find y such thar $P(-y < X < y) = 0.925$

Again, we can do these using the table, but it is a real pain. Fortunately, in R we have qnorm() to solve these.

```
     # find y such that P( X < y ) = 0.23
qnorm( 0.23 )
     # find y such that P( X > y ) = 0.075
qnorm( 1-0.075 )   # this is the ugly old way
qnorm( 0.075, lower.tail = FALSE )  # the better way
     # find y such that P( X < -y or X > y ) = 0.156
     # Because this is a symmetric distribution we can
     # just find y such that P( X > y ) = 0.156/2
qnorm( 0.56/2, lower.tail = FALSE)
     # find y such that P( -y < X < y ) = 0.925
     # By symmetry this is y such that P( X > y ) = (1-0.925)/2
qnorm( (1-0.925)/2, lower.tail = FALSE )
```

```
>      # find y such that P( X < y ) = 0.23
> qnorm( 0.23 )
[1] -0.7388468
>      # find y such that P( X > y ) = 0.075
> qnorm( 1-0.075 )   # this is the ugly old way
[1] 1.439531
> qnorm( 0.075, lower.tail = FALSE )  # the better way
[1] 1.439531
>      # find y such that P( X < -y or X > y ) = 0.156
>      # Because this is a symmetric distribution we can
>      # just find y such that P( X > y ) = 0.156/2
> qnorm( 0.56/2, lower.tail = FALSE)
[1] 0.5828415
>      # find y such that P( -y < X < y ) = 0.925
>      # By symmetry this is y such that P( X > y ) = (1-0.925)/2
> qnorm( (1-0.925)/2, lower.tail = FALSE )
[1] 1.780464
```

This is all well and good but it only takes care of the situation where we have a normal distribution with $\mu=0$ and $\sigma=1$. What do we do in the case where we have a normal distribution with $\mu=76.5303$ and $\sigma=12.04146$? We have only one Normal Distribution table and, at least so far, both pnorm() and qnorm() deal only with our special case.

We will pick up on that idea in Part 2 of this topic.